# Forkable Strings are Rare 

Alexander Russell ${ }^{1}$, Cristopher Moore ${ }^{2}$, Aggelos Kiayias ${ }^{3}$, and Saad Quader ${ }^{1}$<br>${ }^{1}$ University of Connecticut<br>${ }^{2}$ University of Edinburgh<br>${ }^{3}$ 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 $w \in\{0,1\}^{*}$,

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

Observe that $\lambda(w) \geq 0$ 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 $\operatorname{Pr}[w$ 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 $\operatorname{Pr}\left[w_{i}=1\right]=(1-\epsilon) / 2$,

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

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

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

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 $t, \mathbb{E}\left[X_{t+1} \mid X_{0}, \ldots, X_{t}\right] \leq X_{t}$ and $\left|X_{t+1}-X_{t}\right| \leq c$ for some constant $c$. Then for every $\Lambda \geq 0$

$$
\operatorname{Pr}\left[X_{n}-X_{0} \geq \Lambda\right] \leq \exp \left(-\frac{\Lambda^{2}}{2 n c^{2}}\right)
$$

Proof of Bound 1$]$ Let $w_{1}, w_{2}, \ldots$ be a sequence of independent random variables so that $\operatorname{Pr}\left[w_{i}=1\right]=(1-\epsilon) / 2$ 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 $\mathbb{E}\left[W_{t}\right]=-\epsilon$.

Define $\lambda_{t}=\lambda\left(w_{1} \ldots w_{t}\right)$ and $\mu_{t}=\mu\left(w_{1} \ldots w_{t}\right)$ to be the components of the generalized margin for the string $w_{1} \ldots w_{t}$. The analysis will rely on the ancillary random variables $\bar{\mu}_{t}=\min \left(0, \mu_{t}\right)$. Observe that $\operatorname{Pr}[w$ forkable $]=$ $\operatorname{Pr}[\mu(w) \geq 0]=\operatorname{Pr}\left[\bar{\mu}_{n}=0\right]$, so we may focus on the event that $\bar{\mu}_{n}=0$. As an additional preparatory step, define the constant $\alpha=(1+\epsilon) /(2 \epsilon) \geq 1$ and define the random variables $\Phi_{t} \in \mathbb{R}$ by the inner product

$$
\Phi_{t}=\left(\lambda_{t}, \bar{\mu}_{t}\right) \cdot\binom{1}{\alpha}=\lambda_{t}+\alpha \bar{\mu}_{t}
$$

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 $\alpha \bar{\mu}_{n} \leq \lambda_{n}+\alpha \bar{\mu}_{n}=\Phi_{n}$, this implies $\bar{\mu}_{n}<0$, as desired.

Let $\Delta_{t}=\Phi_{t}-\Phi_{t-1}$; we observe that-conditioned on any fixed value $(\lambda, \mu)$ for $\left(\lambda_{t}, \mu_{t}\right)$-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 $\bar{\mu}_{t+1}=\bar{\mu}+W_{t+1}$, where $\bar{\mu}=\max (0, \mu)$; then $\Delta_{t+1}=(1+\alpha) W_{t+1}$ and $\mathbb{E}\left[\Delta_{t+1}\right]=-(1+\alpha) \epsilon \leq-\epsilon$. When $\lambda>0$ and $\mu \geq 0, \lambda_{t+1}=\lambda+W_{t+1}$ but $\bar{\mu}_{t+1}=\bar{\mu}$ so that $\Delta_{t+1}=W_{t+1}$ and $\mathbb{E}\left[\Delta_{t+1}\right]=-\epsilon$. Similarly, when $\lambda=0$ and $\mu<0, \bar{\mu}_{t+1}=\bar{\mu}+W_{t+1}$ while $\lambda_{t+1}=\lambda+\max \left(0, W_{t+1}\right)$; we may compute

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

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

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

Thus $\mathbb{E}\left[\Phi_{n}\right]=\mathbb{E}\left[\sum_{i}^{n} \Delta_{i}\right] \leq-\epsilon n$ and we wish to apply Azuma's inequality to conclude that $\operatorname{Pr}\left[\Phi_{n} \geq 0\right]$ 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

$$
\mathbb{E}\left[\tilde{\Phi}_{t+1} \mid \tilde{\Phi}_{1}, \ldots, \tilde{\Phi}_{t}\right]=\mathbb{E}\left[\tilde{\Phi}_{t+1} \mid W_{1}, \ldots, W_{t}\right] \leq \tilde{\Phi}_{t}, \quad \tilde{\Delta}_{t} \in[-(1+\alpha)+\epsilon, 1+\alpha+\epsilon]
$$

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

$$
\begin{aligned}
\operatorname{Pr}[w \text { forkable }] & =\operatorname{Pr}\left[\bar{\mu}_{n}=0\right] \leq \operatorname{Pr}\left[\Phi_{n} \geq 0\right]=\operatorname{Pr}\left[\tilde{\Phi}_{n} \geq \epsilon n\right] \\
& \leq \exp \left(-\frac{\epsilon^{2} n^{2}}{2 n(1+\alpha+\epsilon)^{2}}\right)=\exp \left(-\left(\frac{2 \epsilon^{2}}{1+3 \epsilon+2 \epsilon^{2}}\right)^{2} \cdot \frac{n}{2}\right) \leq \exp \left(-\frac{2 \epsilon^{4}}{1+35 \epsilon} \cdot n\right) .
\end{aligned}
$$

We give a more detailed argument that achieves a bound of the form $\exp \left(-\epsilon^{3}(1+O(\epsilon)) n / 2\right)$ (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 $\mathrm{A}(Z)=\sum_{t=0}^{\infty} a_{t} Z^{t}$. When $\mathrm{A}(1)=\sum_{t} a_{t}=1$ 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 $\operatorname{Pr}[A=k]=a_{k}$. If the probability generating functions A and B are associated with the random variables $A$ and $B$, it is easy to check that $\mathrm{A} \cdot \mathrm{B}$ 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 $\sum_{t \leq T} a_{t} \leq \sum_{t \leq T} b_{t}$ for all $T \geq 0$; we write $\mathrm{B} \leq \mathrm{A}$ 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 $\operatorname{Pr}[A \geq T] \geq \operatorname{Pr}[B \geq T]$ for every $T$. If $\mathrm{B}_{1} \leq \mathrm{A}_{1}$ and $\mathrm{B}_{2} \leq \mathrm{A}_{2}$ then $\mathrm{B}_{1} \cdot \mathrm{~B}_{2} \leq \mathrm{A}_{1} \cdot \mathrm{~A}_{2}$ and $\alpha \mathrm{B}_{1}+\beta \mathrm{B}_{2} \leq \alpha \mathrm{A}_{1}+\beta \mathrm{A}_{2}$ (for any $\alpha, \beta \geq 0$ ). Finally, we remark that if $\mathrm{A}(Z)$ is a generating function which
converges as a function of $Z$ for $|Z|<R$, it follows that $\lim _{n \rightarrow \infty} a_{n} R^{n}=0$ and $a_{n}=O\left(R^{-n}\right)$; if A is a probability generating function associated with the random variable $A$ then it follows that $\operatorname{Pr}[A \geq T]=O\left(R^{-T}\right)$.

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 $\operatorname{Pr}\left[w_{t}=1\right]=p$. 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

$$
\mathrm{L}(Z)=\sum_{t=0}^{\infty} \ell_{t} Z^{t}
$$

where $\ell_{t}=\operatorname{Pr}\left[t\right.$ is the last time $\left.\mu_{t}=0\right]$. Controlling the decay of the coefficients $\ell_{t}$ suffices to give a bound on the probability that $w_{1} \ldots w_{n}$ is forkable because

$$
\operatorname{Pr}\left[w_{1} \ldots w_{n} \text { is forkable }\right] \leq 1-\sum_{t=0}^{n-1} \ell_{t}=\sum_{t=n}^{\infty} \ell_{t}
$$

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 $\sum_{t \geq n} \hat{\ell}_{n}$. Observe that if $L \leq \hat{L}$ then the series $\hat{L}$ gives rise to an upper bound on the probability that $w_{1} \ldots w_{n}$ is forkable as $\sum_{t=n}^{\infty} \ell_{t} \leq \sum_{t=n}^{\infty} \hat{\ell}_{t}$.

The coupled random variables $\lambda_{t}$ and $\mu_{t}$ are Markovian in the sense that values $\left(\lambda_{s}, \mu_{s}\right)$ for $s \geq t$ are entirely determined by $\left(\lambda_{t}, \mu_{t}\right)$ and the subsequent values $W_{t+1}, \ldots$ of the underlying variables $W_{i}$. We organize the sequence $\left(\lambda_{0}, \mu_{0}\right),\left(\lambda_{1}, \mu_{1}\right), \ldots$ into "epochs" punctuated by those times $t$ for which $\lambda_{t}=\mu_{t}=0$. With this in mind, we define $\mathrm{M}(Z)=\sum m_{t} Z^{t}$ 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, $\mathrm{M}(Z)$ is not a probability generating function, but rather $\mathrm{M}(1)=1-\epsilon$. It follows that

$$
\begin{equation*}
\mathrm{L}(Z)=\epsilon\left(1+\mathrm{M}(Z)+\mathrm{M}(Z)^{2}+\cdots\right)=\frac{\epsilon}{1-\mathrm{M}(Z)} . \tag{1}
\end{equation*}
$$

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 $\mathrm{L} \leq \hat{L}$ 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, $\mathrm{D}(Z)=q Z+p Z \mathrm{D}(Z)^{2}$, and from this we may conclude

$$
\mathrm{D}(Z)=\frac{1-\sqrt{1-4 p q Z^{2}}}{2 p Z}
$$

We likewise consider the generating function $\mathrm{A}(Z)$ for the ascent stopping time, associated with the first time the walk, starting at 0 , visits 1 : we have $\mathrm{A}(Z)=p Z+q Z \mathrm{~A}(Z)^{2}$ and

$$
\mathrm{A}(Z)=\frac{1-\sqrt{1-4 p q Z^{2}}}{2 q Z}
$$

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 $\mathrm{A}(1)=p / q$.

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 $p+q(p / q)=2 p=1-\epsilon$, 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 $t^{\prime}$ for which $\lambda_{t^{\prime}}=0$ 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 $\mathrm{D}^{t}$, the time to descend $t$ steps. This yields the following generating function $\hat{M}$ which, as described, stochastically dominates $M$ :

$$
\hat{\mathrm{M}}(Z)=p Z \cdot \mathrm{D}(Z)+q Z \cdot \mathrm{D}(Z) \cdot \mathrm{A}(Z \cdot \mathrm{D}(Z)) .
$$

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

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

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

Equivalently,

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

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

$$
\operatorname{Pr}\left[w_{1} \ldots w_{n} \text { is forkable }\right]=\exp \left(-\epsilon^{3}(1+O(\epsilon)) n / 2\right) .
$$

## 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.

