Probability of being left behind and probability of becoming permanent tip in the Tangle v0.2

Bartosz Kuśmierz[*] Alon Gal

IOTA Foundation

April 16, 2018

Abstract

We formalize, analyze and numerically estimate probability that given transaction will be *left behind* and probability that transaction will become permanent tip. Analyzed data are gathered for different values of λ and α. As a byproduct of our study, we provide properties of probability of leaving behind in the limit case of parameters.

1 Introduction - permanent tips and transactions left behind

The goal of this document is to discuss the probability of being left behind (POBLB). In terms of confirmation rate: 1- POBLB equals confirmation rate. We also look at the somewhat related concept of probability of becoming permanent tip (POBPT). The latter is less useful than POBLB, as it does not make a difference to the end user whether a transaction will receive zero or negligible number of confirmations. Before we formalize both concepts we introduce the following:

GHOST particle - particle that moves according to MCMC random walk with $\alpha = \infty$ i.e. particle always hops to transaction with highest cumulative weight. In the case of two or more transactions with the same highest cumulative weight, particle hops to one of them with uniform probability.

Final destination of GHOST particle released at time t - last transaction in random walk of GHOST particle (tip in which particle ends up). We denote it by $G(t)$.

Transaction x **directly approve** transaction y, when there is a link from x to y. One writes $x \rightsquigarrow y$.

Transaction x **reference** transaction z, when there is a sequence of transactions such: $x \rightsquigarrow y \rightsquigarrow ... \rightsquigarrow z$.

We denote set of transactions **not** referenced by transaction x as: $\neg x$.

[*] Author's contact information: `bartosz.kusmierz@iota.org`

1.1 Transactions left behind

POBLB can be intuitively characterized as the probability that a given transaction will not be approved by the "majority" of incoming transactions after a long period of time has passed. It is important to note that even if a transaction receives direct approval at some point in time, it can still be left behind. For example it might have only a single transaction approving it, which itself remains a permanent tip. Such a "subbranch", composed of two transactions in this example, would be orphaned, and ignored by the Tangle consensus.

Although these definitions are easy to understand, they are somewhat ambiguous. One can ask: what does it means to be approved by the majority of incoming transactions? How long is a long period of time? To avoid giving answers to these questions, we propose a different, more formal definition, that hopefully captures the same idea:

We say that a transaction is left behind when it is not referenced by GHOST particles in the infinite time limit.

A transaction *tx* is "referenced by GHOST particles at the infinite time limit" when, starting from certain point of time, all final destinations of GHOST particles references *tx*.

$$tx \text{ is not left behind} \Leftrightarrow \exists T : \forall t > T \;\; G(t) \text{ reference } tx.$$

1.2 Permanent tips

POBPT is the probability that a given transaction will not be approved by any other transaction, or remain a *permanent tip*. This quantity is less interesting from the end user perspective; however, it is easier to grasp than POBLB and also easier to estimate. Moreover, it requires almost no additional computational cost. Obviously POBPT $\leq$ POBLB.

2 Towards approximation of POBLB and POBPT

The key point in estimation of POBLB and POBPT are observations that:

1) number of tips
2) number of transactions **not** referenced by $G(t)$

both grow linearly with time, for constant λ. We use the slopes of these lines to estimate POBLB and POBPT.

2.1 Linear approximations

Previous simulations [1, 2] revealed that the number of tips $L(t)$ is linear in time. This result is consistent for a wide range of α and λ. What has not been done so far is analysis of the trend of the number of transactions **not** referenced by $G(t)$ (i.e. $\neg G(t)$). As our simulations show, this quantity also fluctuates around a linear function in time. We present the evolution of both number of tips (green curve) and number of $\neg G(t)$ (red curve) in a series of plots for different α and λ (figs. 1, 2, 3 and 4). More plots are available in the **Appendix A**.

One observes that for $\alpha = 0$, $\neg G(t)$ reaches stable value after certain adaptation period. For small values of λ equilibrium is reached very quickly, but for larger values period of adaptation is visible (roughly first 2000 transactions in the fig. 2, first 500 transactions in the fig. 17). During estimation of POBLB we assured that data were gathered only after adaptation period is over.

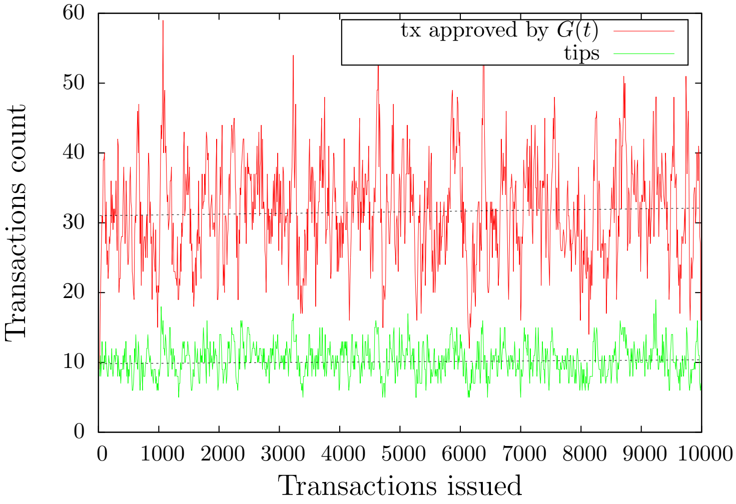

Figure 1: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 0$, $\lambda = 5$. Dashed black lines are trend lines.

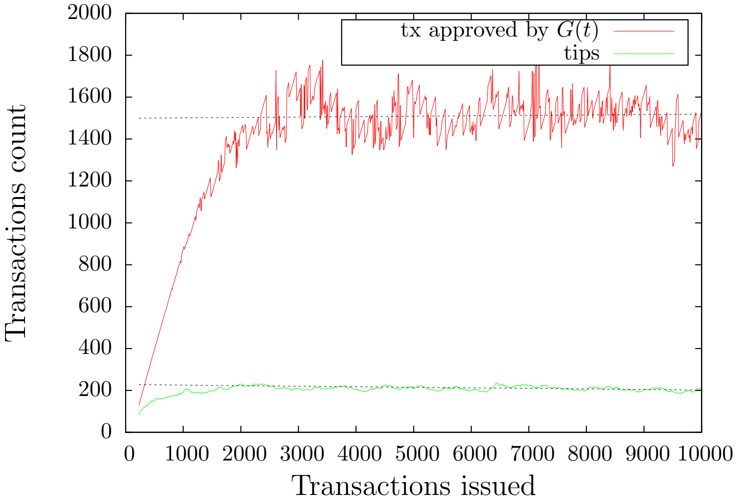

Figure 2: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 0$, $\lambda = 100$. Dashed black lines are trend lines (data are fitted after quantities reach phase of linear growth).

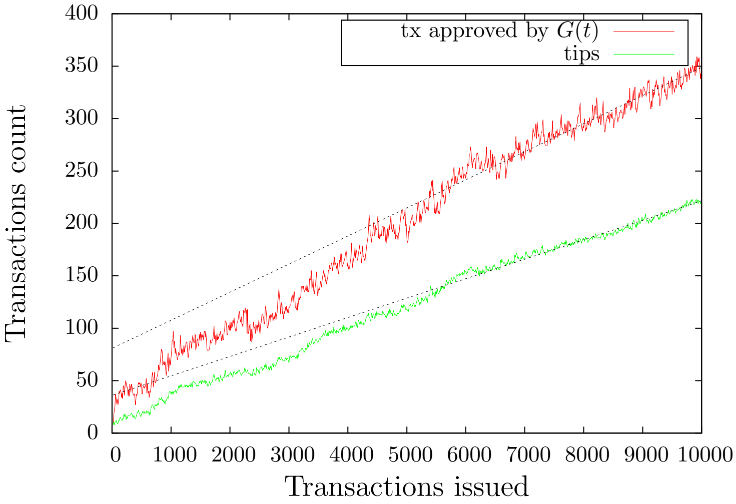

Figure 3: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 0.1$, $\lambda = 5$. Dashed black lines are trend lines.

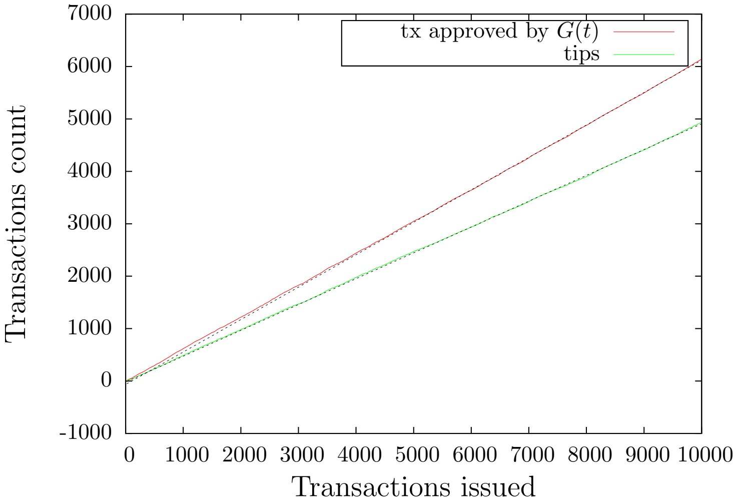

Figure 4: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 1$, $\lambda = 5$. Dashed black lines are trend lines.

As we observe, both tip count and number of transactions not referenced by $G(t)$ fluctuates around trend line. Slope of best linear fit heavily depends on the value of the parameter α. Dependence on λ can be observed as well, however for small α POBLB seems to little affected by changes of λ for examined values of parameters.

Under the assumption that observed linear trend continues, one concludes that roughly the same number of transaction is left behind in each unit of time. Then one can obtain POBLB with the flowing expression

$$\mathrm{POBLB}(\alpha, \lambda) = \lim_{t \to \infty} \frac{\text{transactions } \textbf{not} \text{ referenced by } G(t)}{\text{number of transactions issued until time } t} \tag{1}$$

for constant λ denominator can be validly approximated with simple: $\lambda \cdot t$.

Since we already used the fact that linear trend is observed, we can further estimate POBLB with a slope of lines of best fit presented in the figures 1- 4.

What is worth noticing, final destination of arbitrary GHOST particle does not approve: a) transactions left behind, b) recently announced transactions, positioned "unluckily" with respect to final destination. Transactions from the category b) will be eventually referenced by GHOST particle released at subsequent time. We stress that transactions from this category do not influence our estimation method, as information about them is confined within intercept of line and does not affect the slope.

Similarly POBPT is estimated from the slope of line of tip counts.

Definition based on slopes of linear trend is better than obvious count of transactions $\neg G(t)$ to all of the transactions as it takes into account only increase of abandoned transactions per unit of time.

2.2 POBLB and POBPT - results for fixed λ

Results of the simulation for fixed λ and varied α are presented in the plots 5,6 and 7. For very small α POBLB and POBPT is close to zero and unsurprisingly increases with α. Both quantities grow to the certain limit value. This limit value is not 1, as even for $\alpha = \infty$ certain number of transaction are going to be approved. As we observe limit value grows with λ, we discuss it in more detail in the subsection 3.1.

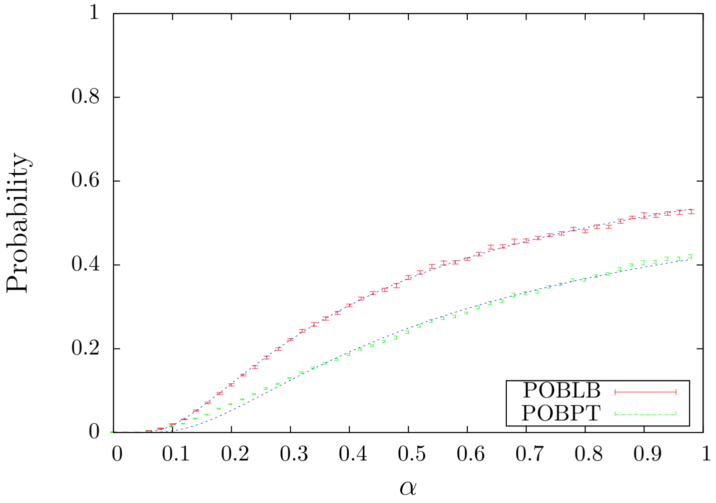

Figure 5: POBLB (red points) and POBPT (green points) as the functions of α for $\lambda = 4$. Blue dashed curves are best fits with functions of type $\exp(a/\alpha + b)$.

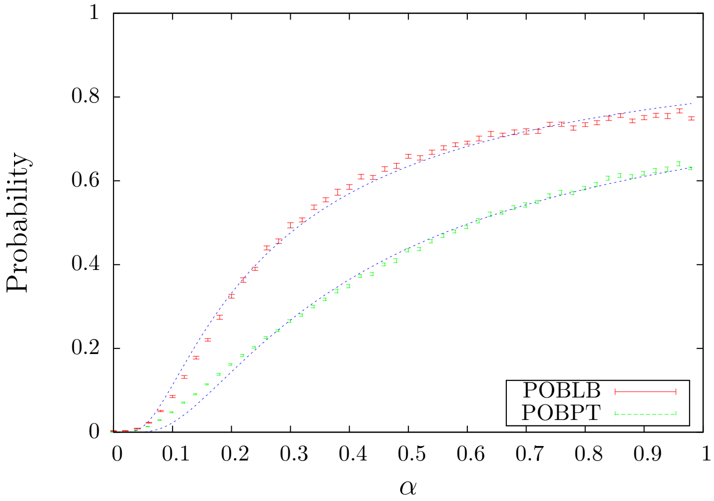

Figure 6: POBLB (red points) and POBPT (green points) as the functions of α for $\lambda = 8$. Blue dashed curves are best fits with functions of type $\exp(a/\alpha + b)$.

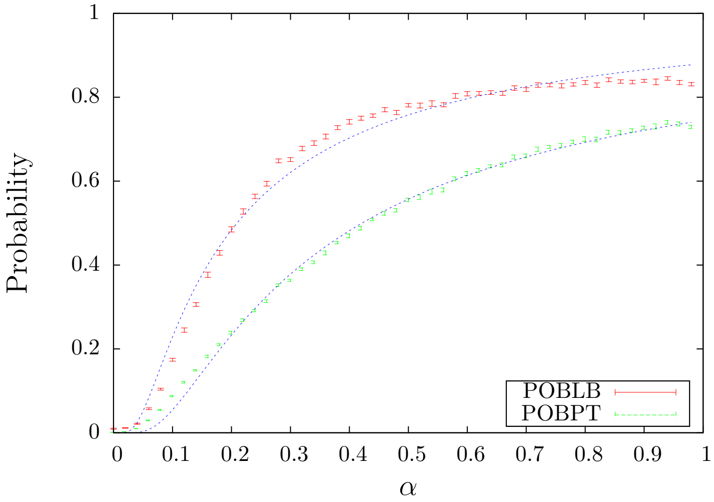

Figure 7: POBLB (red points) and POBPT (green points) as the functions of α for $\lambda = 12$. Blue, dashed curves are best fit with functions of type $\exp(a/\alpha + b)$.

We find it fruitful to plot POBLB as a function of $1/\alpha$ (analogue of temperature $T = 1/\alpha$) on the logarithmic scale. Then the curves "straighten out" - POBLB and POBPT are up to a good approximation linear. This is especially true in the high α regime: see fig 8. This suggests that these quantities can be well approximated by a function of a form

$$a \cdot \exp\left(\frac{b}{\alpha}\right)$$

for some a, b. However we stress that this formula is only a rough approximation and the fit is not consistently effective on the whole domain. We acknowledge that regimes of small and large $1/\alpha$

have slightly different slopes on log scale. Hence, if one wants to approximate them with function of this type, one should consider using different fits for regimes of large and small α.

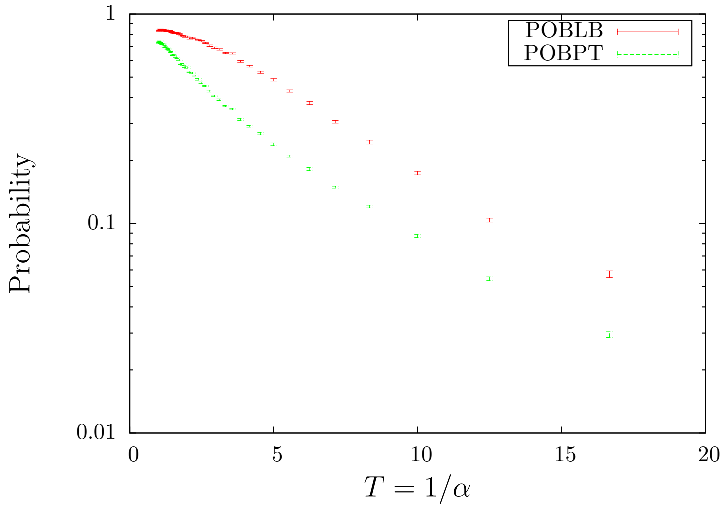

Figure 8: POBLB (red points) and POBPT (green points) as the functions of α for $\lambda = 12$. Similar picture can be observed for other values of λ as well.

3 estimations of POBLB and POBPT

We plotted 2D maps of POBLB and POBPT as a function of α and λ on the figs. 10, 11,12 and 13. A single run of the simulation involved 5000 transactions and each point on the plot is averaged over 49 simulation runs. A new transaction starts a particle walker from the genesis, and performs the MCMC random walk towards a tip. The first tip to which this particle walker arrives is chosen as the first approval site. Then, to find the second approval site, this process is repeated until it finds a new tip or has arrived at the same tip 10 consecutive times, in which case it will choose as the second approval site one of the two transactions directly approved by this tip.

What is worth noticing both probabilities are increasing with λ. **Conclusion from those plots is as follows: If one wants to keep percentage of transaction left behind to be constant, then as λ grows α has to be decreased.**

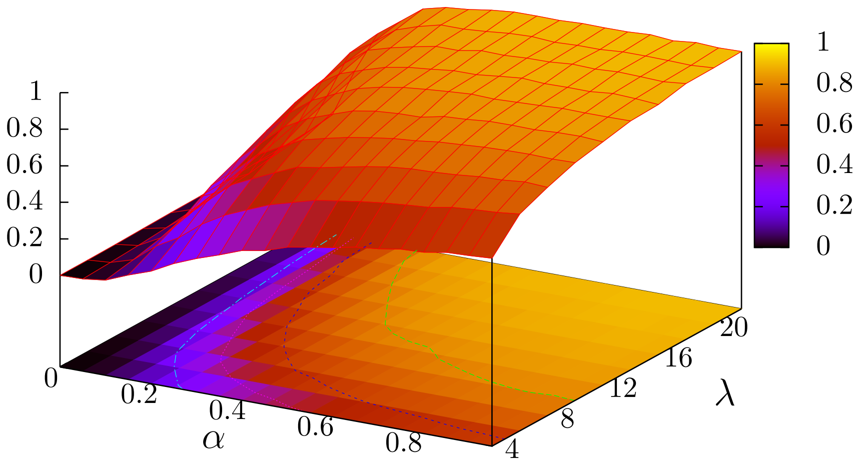

Figure 9: POBLB as a function of α and λ for $\alpha \in [0, 1]$. Plot with contour lines for 0.2 (cyan), 0.4 (pink), 0.6 (blue) and 0.8 (green).

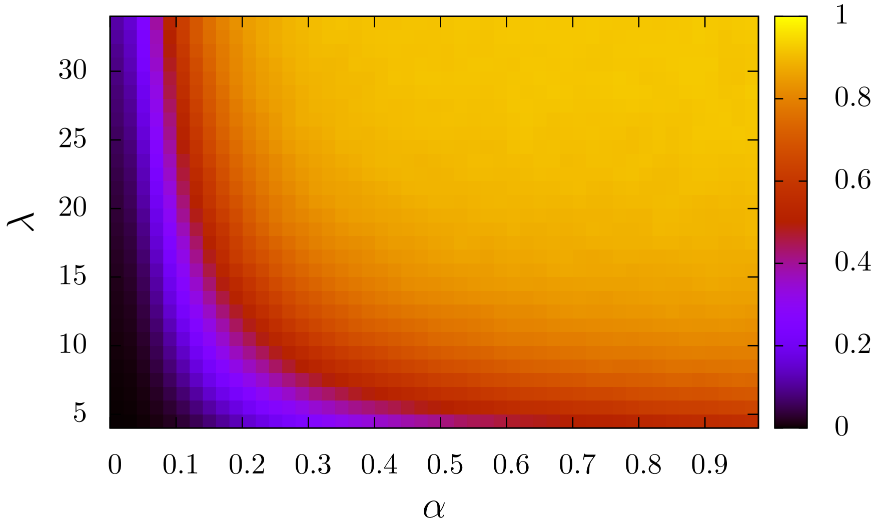

Figure 10: POBLB as a function of α and λ for $\alpha \in [0, 1]$.

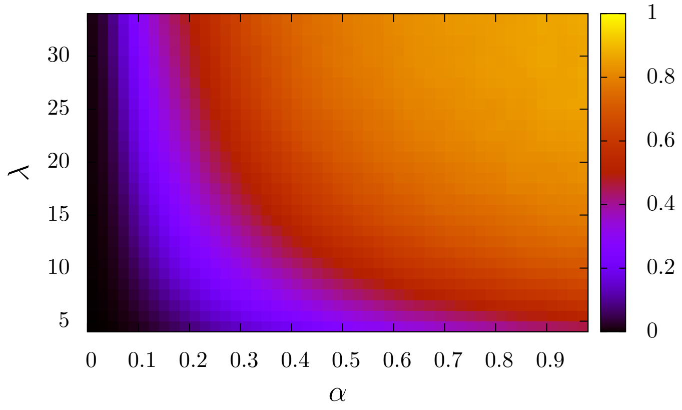

Figure 11: POBPT as a function of α and λ for $\alpha \in [0, 1]$.

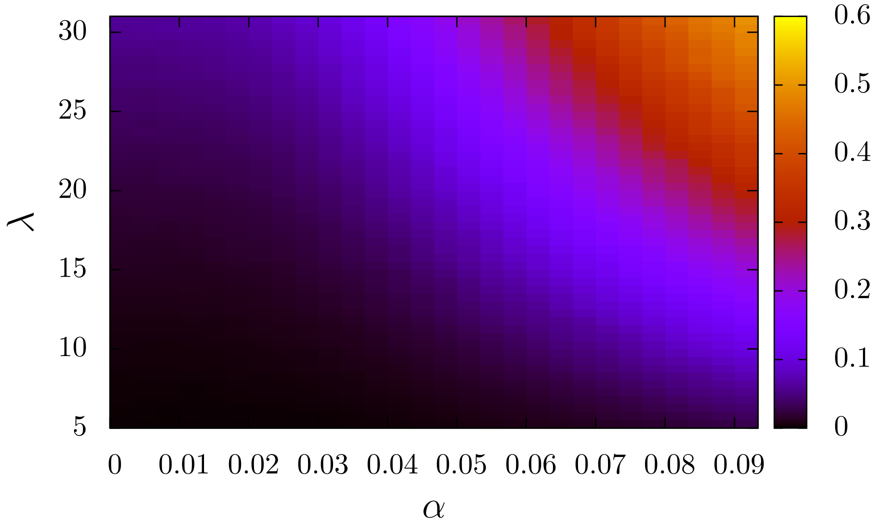

Figure 12: POBLB as a function of α and λ for $\alpha \in [0, 0.1]$.

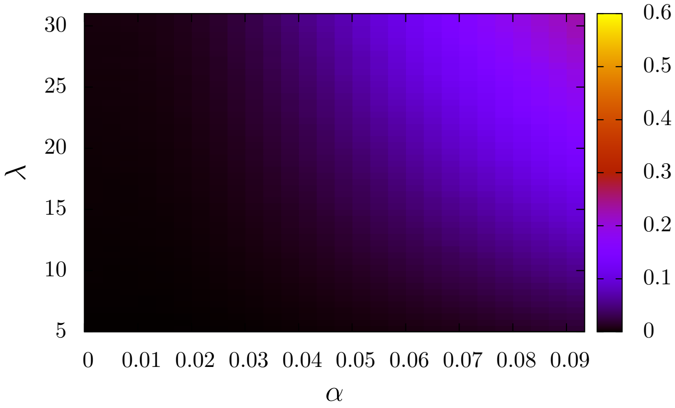

Figure 13: POBPT as a function of α and λ for $\alpha \in [0, 0.1]$.

Search for approximate formula for POBLB as a function of α and λ led us to plot it on a logarithmic scale as a function of $T = 1/\alpha$ and $1/\lambda$. Then POBLB surface seems to be approximately a plane (see fig. 14), especially for smaller $1/\alpha$ and $1/\lambda$. This is not a perfect fit, one can easily observe that contour lines are not exactly straight lines, as it would be in the case of plane and reveal more "hyperboloid-like" shape. Nevertheless such fit with plane on logscale would require only three parameters. Then POBLB would be of a form

$$\mathrm{POBLB}(\alpha, \lambda) = a \cdot \exp\left(\frac{b}{\alpha} + \frac{c}{\lambda}\right) = a \cdot \exp\left(\frac{b}{\alpha}\right) \cdot \exp\left(\frac{c}{\lambda}\right) \tag{2}$$

for some a, b, c.

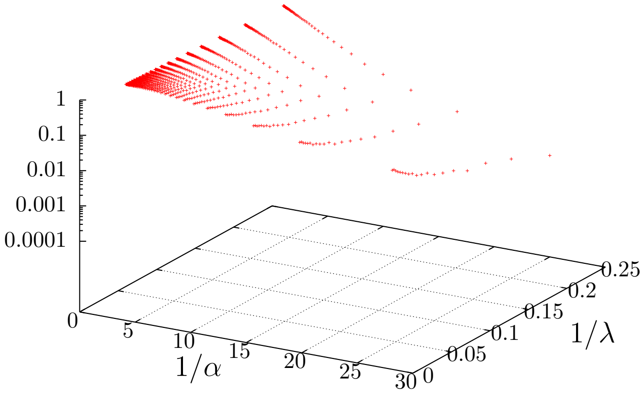

Figure 14: POBLB as a function of $1/\alpha$ and $1/\lambda$, logarithmic scale.

Under the assumption that equation 2 is correct lines of constant POBLB are determined by

$$\log(\mathrm{POBLB}) - \log(a) = \frac{b}{\alpha} + \frac{c}{\lambda},$$

or (for fixed POBLB and λ)

$$\alpha = \frac{b}{\lambda\left(\log(\mathrm{POBLB}) - \log(a)\right) - c}. \tag{3}$$

What is worth stressing equation 3 gives: $\lambda \to \infty \Rightarrow \alpha = 0$. However, we stress that fit with function of a type (2) is debatable in limit cases.

3.1 POBLB for $\alpha = \infty$

Results presented in the fig. 10 suggest that even for small λ, POBLB reach its highest value (saturates) for α around and below 1. For larger λ limit value is obtained even faster. Thus for all purposes MCMC random walk with $\alpha = 10$ should be the same as $\alpha = \infty$. We use it as an approximation of $\alpha = \infty$. We stress that for $\alpha = 10$ probability that particle will not hop to the transaction with highest cumulative weight is smaller or equal to $1 : \exp(10 \cdot 1) \approx 1 : 22000$ (smallest difference in cumulative weight equals 1). We present POBLB as a function of λ for $\alpha = 10$ in a fig. 15. We fit data with curve of a type

$$\mathrm{POBLB}(\lambda) = a + \frac{b}{\lambda + c}.$$

Values of parameters are $a = 0.9907 \pm 0.0019, b = -1.47 \pm 0.03, c = 1.061 \pm 0.046$. This suggests that number of transactions “approved by GHOST particles at infinity” (particles regarded as not left behind) decays like $1/\lambda$. We stress that parameter a up to the uncertainty equals 1.

When $\alpha = \infty$ and particles are deep in the tangle, they stick to the “path” composed out of transactions with highest cumulative weight. Such “path” is thin, however it may branch as there can be more than one transaction with highest weight. When particle is in near the end of the tangle and more than one tip is available then new incoming transaction approves two of them with equal probability (as all of them have the same cumulative weight equal to 1). But as quickly as single transaction that approve any of the tips is anunced, all new particles will pass throught this (former) tip and it will most likelly be a newest part of the “path”. Occasionally, more than one transactions can be anounced at the same time, but this event should be equally probable during simulation and would affect only coefiecient b of fit

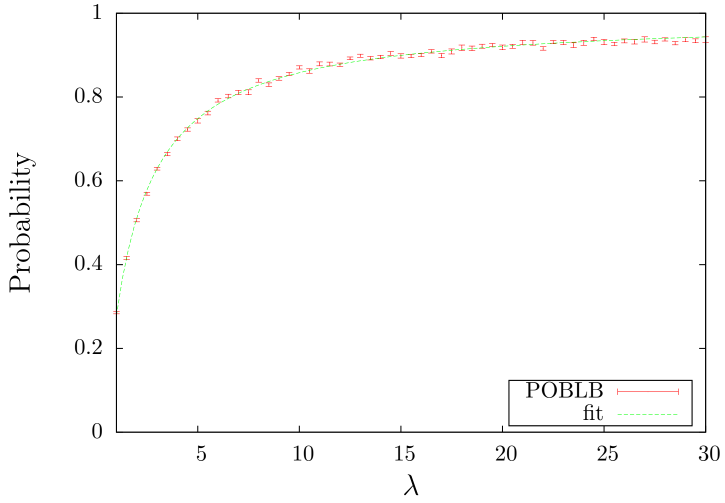

Figure 15: POBLB for $\alpha = 10$.

4 Conclusions

We have derived a quantitative estimation of the POBLB, which can directly be translated into parameters for running IOTA nodes.

Assuming we know how much time PoW takes on average (h), and how many transactions come in per minute (lambda), we can decide what confirmation rate we are interested in (POBLB) and set alpha appropriately, using the information in figure 10.

On the theory side, we have confirmed the following:

1. We confirmed that the number of abandoned transactions follows a linear trend, similarly to $L(t)$.

2. POBLB and POBPT follow similar trends, and either is a reasonable approximation for computing the confirmation rate.

3. We have derived a first attempt at an equation of state of the Tangle, using POBLB: $\frac{b}{\alpha} + \frac{c}{\lambda} = \log(\text{POBLB}) - a$. This result may be compared against future analytical work, as well as real IOTA data.

4. Qualitatively, if we fix POBLB to a desired level, an increase in λ means α must be lowered.

5. TBD: summarize behavior in limit cases of alpha and lambda 0 or ∞ . I didn't get the bottom line.

5 Further work

1. Looking at figure 10, and following curves of constant POBLB, we see vertical asymptotes. It is interesting to ask whether all curves asymptote at $\alpha = 0$, or if rather there a minimum α value for each POBLB? The latter option, if true, would be powerful for IOTA. It implies that we can set α independently of λ, rather than adjust it dynamically.

2. We seem to have a good measure for what α should be, based on our desired confirmation rate. The natural next question is how to quantify the security motivation, that will serve as the counterpart. An ideal α value should capture a compromise between the system's throughput and security.

A Additional plots

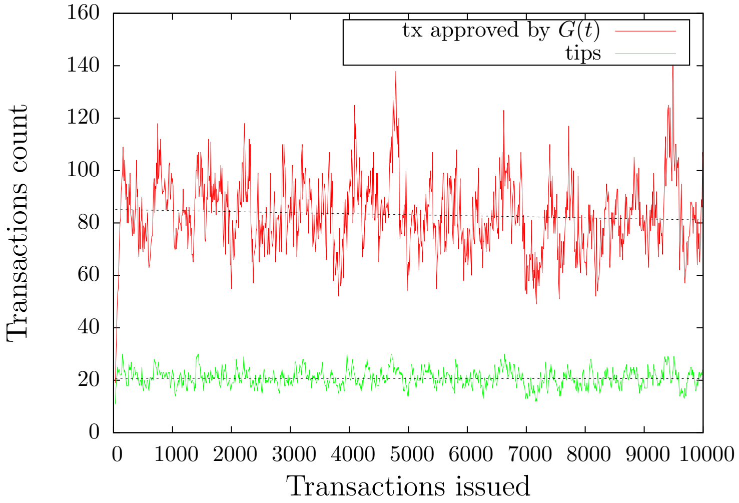

Figure 16: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 0$, $\lambda = 10$. Dashed black lines are trend lines.

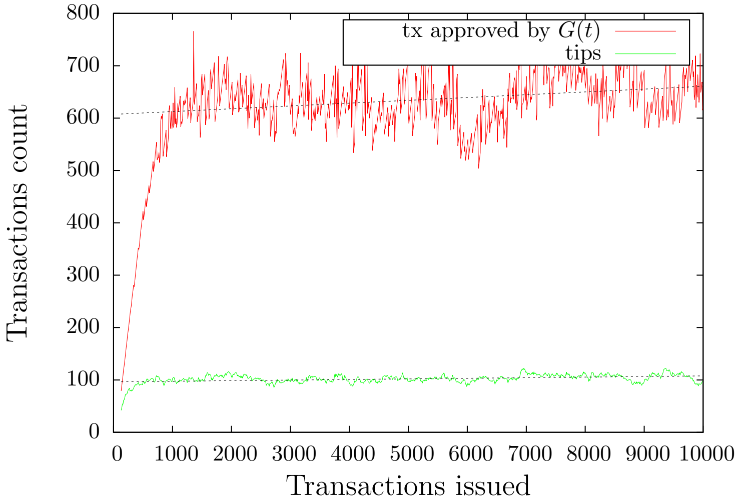

Figure 17: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 0$, $\lambda = 50$. Dashed black lines are trend lines (data are fitted after quantities reach phase of linear growth).

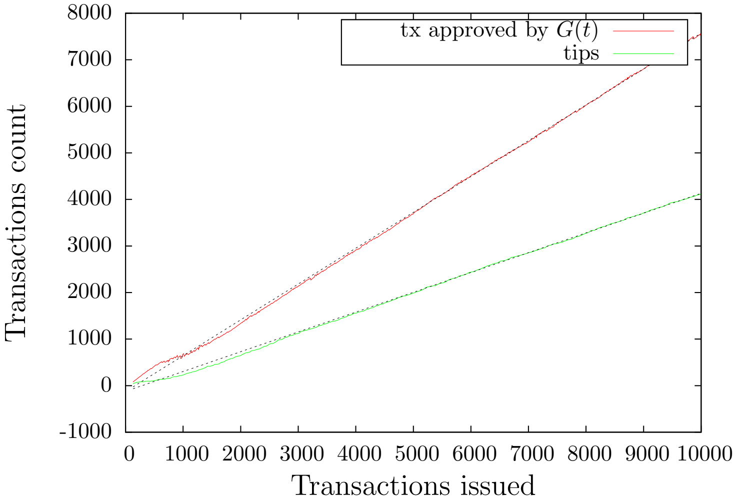

Figure 18: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 0.1$, $\lambda = 50$. Dashed black lines are trend lines.

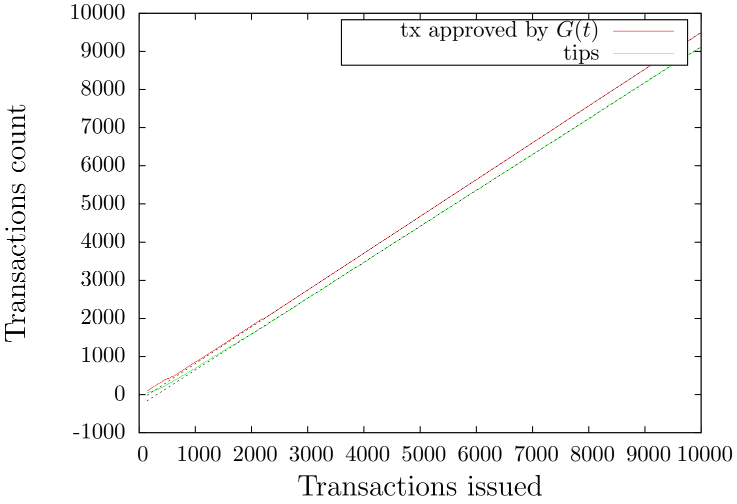

Figure 19: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 1$, $\lambda = 50$. Dashed black lines are trend lines.

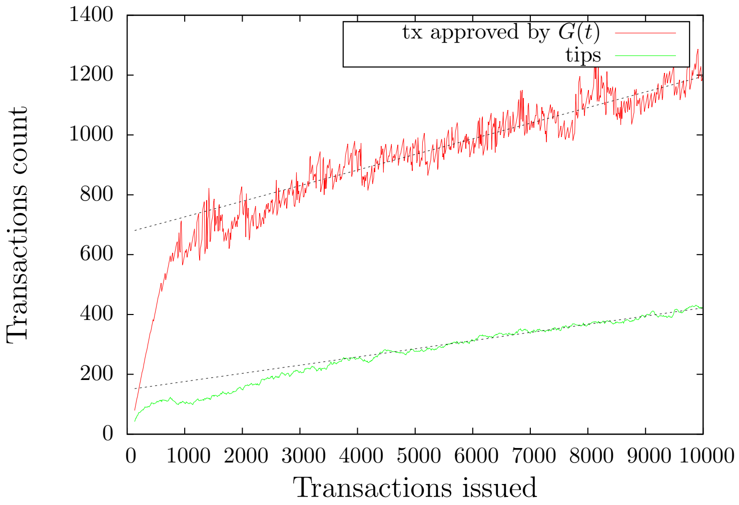

Figure 20: Number of transactions not approved by $G(t)$ (red curve) and number of tips (green curve) for $\alpha = 0.025$, $\lambda = 50$. Dashed black lines are trend lines.

References

[1] B. Kuśmierz, “The first glance at the simulation of the Tangle: discrete model”, http://iota.org/simulation_tangle-preview.pdf (2017)

[2] B. Kuśmierz, P. Staupe and A. Gal, “Extracting Tangle Properties in Continuous Time via Large-Scale Simulations” (working paper).