Maypoles: Lightning Striking Twice

Clara Shikhelman
Chaincode Labs
clara.shikhelman@gmail.com

Abstract

The Lightning Network (LN) is a second layer solution built on top of Bitcoin, aimed to solve Bitcoin's long transaction waiting times and high transaction fees. Empirical and theoretical studies show that the LN is tending towards the hub and spoke network topology. In this topology most of the nodes, the spokes, open a single channel to one of the few well-connected nodes, the hubs. This topology is known to be prone to failures, attacks, and privacy issues. In this work we introduce the *Maypoles* protocol in which most nodes open two channels instead of one. We show that this protocol benefits the network significantly by enhancing its stability, privacy, and resilience to attacks. We also examine the economic incentives of nodes to take part in Maypoles.

1 Introduction

The limited throughput is a challenge shared by all Proof of Work (PoW) blockchains, such as Bitcoin. By design, these blockchains limit the expected number of transaction per second, as making this number larger comes at the expense of security.

A direct consequence of this is an auction-like dynamic for the limited space available, driving the transaction fees up. High fees limit the usability of the blockchain, and narrow its adoption.

Another limit of PoW blockchains is the long waiting time needed for a transaction to be considered secure. In PoW blockchains, the longer one waits, the more secure the transaction is. The expected time a transaction waits until it is approved varies between blockchains, usually being somewhere between a few minutes to an hour. Even a minute is an unacceptable waiting time for retail transactions and other use-cases. Similar to the throughput problem, this cannot be changed without serious security risks.

Second layer protocols are solutions to the above challenges. Generally speaking, these protocols aggregate transactions and send a summary to the blockchain as seldom as possible. By doing this, these protocols reduce congestion on the blockchains and offer cheaper and quicker transactions.

In this paper, we focus on Bitcoin's most predominant second layer solution, the Lightning Network (LN) introduced in [24]. The results presented here are

relevant in the wider context of payment channel networks that work on the same principles. See [14] for an overview.

The basic building block of the LN is a channel. Alice and Bob open a channel by sending a Bitcoin transaction. This transaction locks a given amount of bitcoin inside the channel. Once the channel is open, Alice and Bob can freely transact with each other by shifting the amount of funds owned by each party in the channel. These transactions happen immediately, and do not entail any fees. At any point, Alice and Bib can close the channel by sending another Bitcoin transaction and get their respective funds back on the blockchain.

Furthermore, if Bob also has a channel with Carol, Alice can transact with Carol through Bob, if Bob allows this. There is no theoretical limit[1] on the length of the chain of payments. So, Alice can pay Zoey through Bob, Charlie, Donna, and so on, as long as all the parties on the way from Alice to Zoey agree to this. Parties along the route may charge a fee. We give further details on payments over the LN in Appendix A.

Several studies have shown the tendency of the LN towards the *hub and spoke* topology. This topology can be described as follows. There is a small number of nodes with many channels, these nodes are called *hubs*. The rest of the nodes are connected to a single hub. A node connected with a single channel to a hub is called the hub's *spoke*. See Figure 1 for an illustration.

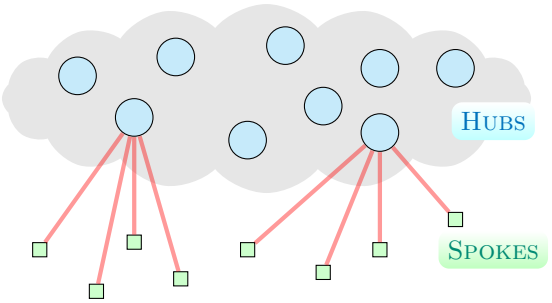

Figure 1: Hub and spoke topology. Each hub, a blue circle, has many spokes. Each spoke, a green squares, is connected to a single hub.

Hub and spoke networks suffer from stability, security, and privacy issues. These issues arise both from the lack of guarantees provided by the connections between the hubs and from the fact that the spokes have only a single channel.

In this paper, we propose a solution called *Maypoles*. This solution builds on top of the hub and spoke topology. In Maypoles, each spoke opens two channels instead of one. The first channel to a hub of its choice, called *the main hub*, through which it plans to move most of its transactions. The second channel to a hub chosen by the spoke's main hub, this hub is called *the secondary hub*. The main hub selects the secondary hub randomly. The full details of the protocol are given in Section 2.2.

[1] In practice, there is a limit of 20 hops in some implementations

We show that by adding channels as prescribed by Maypoles, we are able to make the network exhibit a healthy topology that is less prone to attacks and breakdowns due to malfunction, and to enhances privacy. We also show that it is in the best interest of the spokes to open these two channels, as this allows them to send payments over the LN without down-times and save on channel rebalancing fees.

1.1 Our Contribution

Topological improvements In this paper we make the first attempt, to the best of our knowledge, to push back against the tendency of the LN to collapse to a centralized structure such as the hub and spoke topology. While understanding that currently most nodes open a single channel to a well-connected hub, we offer to build on top of this to improve the network's structure.

We prove theoretical results, with no assumptions on the structure of the underlying LN, and then show the results of simulations on top of current LN data. The simulations show that the average case over the current LN is even better than the guarantees given by the theoretical results.

Local change for global impact In Maypoles the benefits to the network do not rely on a coordinated effort of all nodes. Each hub chooses its own random subset of secondary hubs, and each node chooses its main hub independently of other nodes in the network. We show that local independent decisions can improve global properties of the network.

Economic incentives To motivate improvements to the LN, we use recent results on the economy of transactions fees in Bitcoin and of LN channels. By examining economic incentives, we create a win-win situation where both the network and the spokes benefit from the protocol. The economic assumption made in proving the incentive compatibility are modest and backed by data.

2 Maypoles

2.1 The Model

The starting point we wish to improve is the hub and spoke topology. We assume that the set of nodes is $V = H \cup S$, where H is the set of hubs and S is the set of spokes. Each spoke $s \in S$ is connected to a single hub by exactly one channel. We assume that each hub $h \in H$ has at least 4 spokes. We also assume that there is some global parameter k all the hubs use for the protocol[2].

We assume that we are in the steady state of the hub and spoke model, that is, each spoke s chose the hub h_s, as it wishes to send and receive all of its transactions through h_s. When the channel between s and h_s becomes

[2]We look at a fixed k to keep the analysis simple, the same results will hold if each hub chooses its own value for this parameter, and we take k to be the smallest of them

unbalanced, that is, it cannot support any more transactions that the spoke wishes to send or receive, the spoke will rebalance the channel. For example, if all of the balance of s in the LN is 0 and s wishes to keep sending transactions, s will move funds from the blockchain to the LN. We give further details on rebalancing strategies in Section 4.1.

We also assume that the fee of a Bitcoin transaction is correlated to the urgency of the transaction, that is, at any moment in time there is a fixed fee that will guarantee with high probability that the transaction appears in the next block. A smaller fee will guarantee that the transaction will appear in the next 6 block, and so on. We also assume that the cost of keeping a channel that is always available for sending or receiving, is a monotone decreasing function of the cost of rebalancing the channel.

2.2 The Protocol

Maypoles is built on top of a network with a hub and spoke topology. In Maypoles, each spoke opens two channels instead of one. From the spoke's viewpoint, it opens a main channel to a hub of their choice. The hub then recommends a hub for the spoke to open their secondary channel to.

To recommend a secondary channel to each spoke, every hub does the following protocol (independently of the other hubs)

(i) Initiate L as an empty list, and let l denote the size of L throughout the protocol.

(ii) Choose k hubs uniformly at random. Add them to L.

(iii) Notify the hubs that they are chosen, and add any hub that chose you to the list L.

(iv) Break the spokes into $l(= |L|)$ sets of size as even as possible.

(v) Instruct the spokes in set i to open their secondary channels to the i'th hub in L.

Notice that there is a symmetry between choosing a hub in step ((ii)) and being chosen by a hub in step (3), that is, there are secondary channels between the spokes of hubs h_0 and h_1 if either h_0 chose h_1 or h_1 chose h_0.

2.3 Overview of the Benefits

If the hub h_0 chose the hub h_1 (or vise versa) then the result will be spokes that are connected to both of them (see Figure 2). This means that transactions can go between h_0 and h_1, creating a new connection. This connection can be used if need be.

For example, in cases of failure in the network, different nodes can use this connection to route their transactions. Spokes can also choose to use this connection to enhance their privacy by occasionally using the secondary channel to

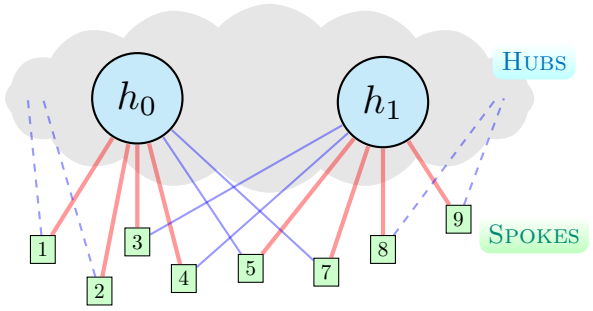

Figure 2: Hubs and spokes in Maypoles. The red thick edges represent the spokes' main channels, and the blue thin edges represent the secondary channels

transact. The main theorem concerning the network topology shows that Maypoles guarantees various network properties. Simulations based on the current LN show that these improvements out perform the theoretical guarantees in the average case. Further details are given in Section 3.

We also make sure that the spokes do not lose money when opening the secondary channel. As the cost of the LN channel strongly depends on the cost of on-chain transactions, we can use the fact that delayed on-chain transaction have lower fees for the benefit of the spokes. Further details are given in Section 4.

3 Maypoles and Network Topology

3.1 Network Topology

The topology of channels in the LN has a crucial effect on the stability, security, and privacy of the network. There are several graph theoretical properties that are of interest when we are considering the topology of the LN. We start by giving an overview of these properties in the context of the LN.

Connectivity We say that a network is connected if for every pair of nodes there is a path of edges between them. We say that a network is m-edge (node) connected, if one needs to remove at least m-edge (node) to make it disconnected. A network's edge connectivity is equivalent to the minimum-cut of the network. In the context of the LN, the edges are LN channels, and the nodes are LN nodes that take part in these channels.

Node and edge connectivity are a way to measure the network's stability. High connectivity in the LN ensures that even if several nodes or channels malfunction, the other nodes can continue to transact with each other. See Figure 3 for an example.

Furthermore, high connectivity ensures the existence of many disjoint paths

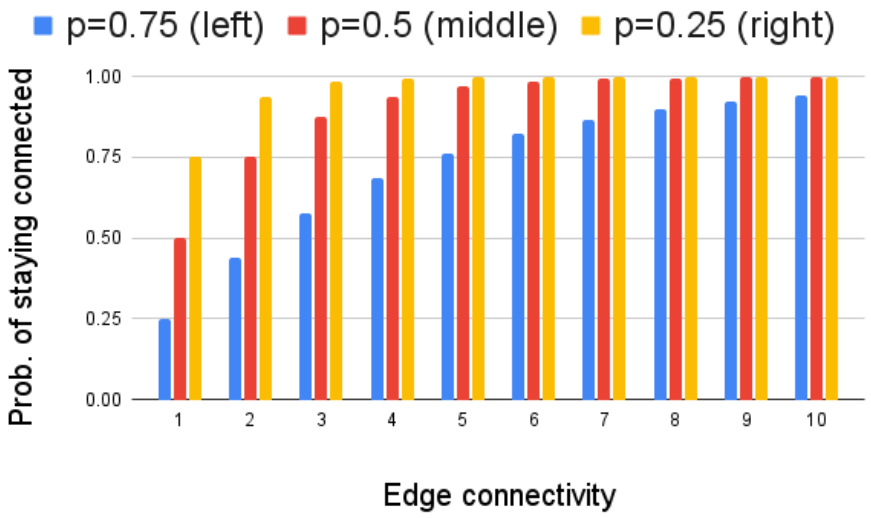

Figure 3: Let p be the probability of an edge failing due to an attack or a malfunction. The above figure shows the probability of the network staying connected as a function of its edge connectivity for different values of p .

going between different nodes in the network. This helps both stability and privacy. For example, if Alice and Zoe don't have a channel between them, and occasionally transact, the nodes along the path that they use can gain information about their transaction habits (see [29] for further details). Having many disjoint paths to choose from helps secure the privacy of both Alice and Zoe by making it significantly more difficult to collect information.

Diameter The distance between two nodes is the length of the shortest path connecting them in the network. The diameter of a graph is the maximum distance between two nodes in the graph. The diameter gives us an upper bound on the distance between nodes, which guarantees the existence of a short path between any pair.

Long paths are a problem in the LN for two reasons. First, any node along the path charges a fee, making the use of long routs expensive. Second, as channels might malfunction or fail to accommodate transactions, each extra channel in the path adds a non-negligible probability of failure.

Currently, the fees in the LN are rather small, but the failure rates are high. Routes mostly fail due to insufficient funds in one of the channels, although nodes going offline and other problems are also an issue. For the payment to go through, we cannot allow even a single failure along the route. This means that the probability of failure grows exponentially with the length of the route. Figure 4 shows how the route successes probability diminishes as the length of

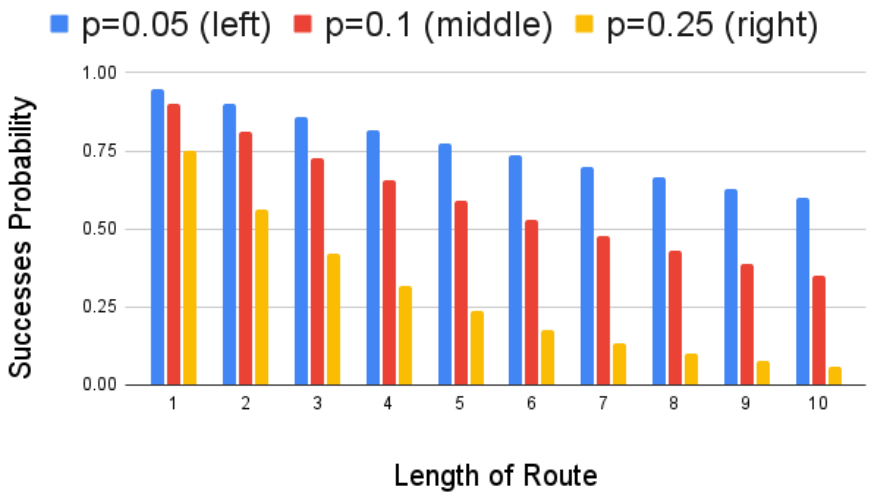

Figure 4: Probability of a route succeeding as a function of the length of the route, assuming the probability of a single channel to fail is p.

the route grows.

A short diameter offers an upper bound on the failure probability for a transaction sent between any pair of nodes. This probability is an important parameter for the usability of the LN.

Expansion A network having high expansion means that any set of nodes has many edges connecting it to the rest of the network. In the LN, having good expansion will ensure that any set of nodes has many channels connecting the set to the other nodes in the network. Having high expansion guarantees many helpful properties. For example, high expansion guarantees that breaking the network into two large components that cannot transact with each other requires removing a very large number of edges (see [18]). This makes a large scale attack extremely difficult to accomplish.

High expansion also guarantees that the network is not centralized around a small set of nodes. This is of particular importance in the LN, as the decentralization helps privacy, censorship resistance and stability.

Betweenness Centrality The betweenness centrality of a node v in a network is the proportion of the shortest paths, between any pair of nodes, that goes through v. Most nodes in the LN choose to route using the shortest path[3] in the network. In light of this, if there is a small set of nodes with a betweenness centrality that is significantly greater than the average, this means that

[3]To be more precise, they choose the cheapest path, yet these are almost always the same

most nodes route through that set. This allows the nodes, especially if they are working together, to learn a vast amount of information about flows in the network. As any form of centralization, it also opens up the network to failures and various attacks.

To make sure that nodes do not that take up a disproportional part of the flow in the network, we look at the values of the betweenness centrality of all the nodes in the network, and take the standard deviation of these values. The smaller the standard deviation is, the better, as this points to the decentralization of the network. As this is a difficult parameter to study formally, we will focus on it only in the simulation part.

The hub and spoke topology cannot guarantee good network parameters. By adding channels as prescribed in Maypoles, we can ensure the network's health. In the next section, we give formal definitions of the above properties, state and prove the main theorem that shows Maypoles indeed improves the topology without making any assumption on the underlying LN. Then we present the results of several simulations where we study the Maypoles protocol over a snapshot of the LN and consider several variations to the protocol.

3.2 The Main Theorem

The theorem we state and prove in this section shows the benefits of Maypoles to the network structure. Before stating and proving our theorem, we need to give formal definitions for the LN under Maypoles and for the desired network properties.

In Maypoles, if a hub, say h_0, has chosen another hub, say h_1, then h_0 and h_1 are connected to each other via a subset of their spokes. This means that h_0 and h_1 can transact with each other, routing through the spokes they share. To study this relation, we define the *secondary network*.

Definition 3.1 (Secondary Network)**.** *The Maypoles protocol with a fixed constant k creates the secondary network $\mathcal{L}_k$. The nodes of $\mathcal{L}_k$ are all the hubs in the LN. There is an edge between hub h_0 and hub h_1 in $\mathcal{L}_k$ if one of the hubs chose the other in stage (2) of the Maypoles protocol, that is, if there are spokes connected simultaneously to h_0 and h_1.*

When considering the influence of Maypoles on the LN, we think of a new network created by adding the edges of $\mathcal{L}_k$ to the existing LN. The LN together with $\mathcal{L}_k$ will have edge and node connectivity greater or equal to the connectivity of $\mathcal{L}_k$. The same holds for expansion. As for the diameter, the distance between nodes in $\mathcal{L}_k$ is an upper bound on the distance in the LN together with $\mathcal{L}_k$, and so the diameter can only become smaller.

To understand how the addition of $\mathcal{L}_k$ improves the topology we first give the formal definitions of the network properties, then we state the theorem, and finally prove it.

Definition 3.2. *Let $G = (V, E)$ be a connected network, where V is the set of nodes, and E is the set of edges.*

- ***Connectivity*** *G is connected if there is a path between any two node. G is m edge (node) connected if it stays connected if we remove any set of $< m$ edges (nodes).*

- ***Diameter*** *For $u, v \in V$, let $d(u, v)$ be the distance between u and v, defined as the length of the shortest path from u to v. The diameter of G is $\max_{u,v \in V} d(u, v)$, that is, the greatest distance between two nodes in G.*

- ***Expansion*** *For a set of nodes $S \subseteq V$ define its outer boundary to be the set of edges that have one end in S and one end outside S, that is, $\partial(S) = \{(u, v) \in E \text{ s.t. } u \in S \text{ and } v \notin S\}$.*

  The edge expansion of G is the smallest ratio between the size of the boundary of a set and the size of the set itself, that is, $\min_{S \subseteq V, 0 < |S| \leq \frac{|V|}{2}} \frac{|\partial(S)|}{|S|}$.

Theorem 3.3. *Let $\mathcal{L}_k$ be Maypoles' secondary network with $k \geq 2$. Then w.h.p.*[4]

(i) *$\mathcal{L}_k$ is k edge connected*

(ii) *$\mathcal{L}_k$ is k node connected*

(iii) *$\mathcal{L}_k$ has an edge expansion of at least $(1 + o(1))k$*

(iv) *$\mathcal{L}_k$ has a diameter of at most $(1 + o(1))\frac{\log n}{k+1}$, where n is the number of nodes.*

Note that the theorem does not assume anything about the underlying LN graph, and only shows properties of $\mathcal{L}_k$. When adding $\mathcal{L}_k$ to the LN, it could happen that the connectivity and expansion would be even better, and the diameter even smaller. When we simulate Maypoles with a snapshot of the LN, we indeed see that the improvement in the average case is even better than is guaranteed by the theorem. Further details on the simulations are given in Section 3.3. The proof of the theorem can be found in Appendix B

3.3 Simulations and Variations

In the previous section, we have focused on theoretical results that allowed us to give bounds on properties guaranteed by the Maypoles protocol, regardless of the structure of the underlying LN. In this section, we simulate the effect of Maypoles on the current topology of the LN. We also discuss and simulate some variations of the Maypoles protocol and their effect on network properties. To simulate Maypoles and its variation, we have used the LN snapshot from [8]. The code of the simulation can be found in [2].

In addition to studying the regular Maypoles protocol, we have also considered the following variations. In the first one, when a hub chooses k hubs to connect to, there is a subset of hubs it prefers over the others. It is natural to

[4] w.h.p. meaning *with high probability*, that is, with probability going to 1 as the number of nodes grows

assume that some hubs would prefer to connect to hubs that are similar to them. The similarity could be in relation to geography, the demography of customers, or other parameters that will make the link between the hubs more advantageous. As we do not have any meta-data on the hubs in the LN snapshot, we chose a random-like heuristic where hubs prefer other hubs if the last digits of their IDs are the same.

In the second variation, some hubs do not cooperate, that is, these hubs do not instruct their spokes to open secondary channels. It could easily happen that a hub would signal that it is interested in taking part in Maypoles, and then fail to perform the protocol. In many cases, hubs are interested in a change, yet will take a very long time before adopting it in practice. To simulate this, each hub chooses whether to cooperate or not by flipping a coin with some fixed success probability. Hubs do not know which hubs cooperate and which don't, and so a cooperating hub might instruct its spokes to connect to a non-cooperating hub. This will help the non-cooperating hubs, but not as much as performing the protocol.

We examine Maypoles and the above variations through two parameters. The first is the edge connectivity of the network. This will tell us both how resilient is the network to failures and attacks, and will show us the number of different routes available between pairs of nodes. If hubs deviate from the uniform choice by preferring some hubs, we expect to see similar results to the uniform case, as there is a variety of edges added for every hub. Because connectivity is very sensitive to local changes, we expect a significant change if a large proportion of hubs do not cooperate.

The second parameter we want to examine is the betweenness centrality in the network. Most nodes in the LN will choose to route through the shortest path available. In any connected network, for any pair of nodes $\{s, t\}$, there is at least one path of minimal length. For a hub h, the betweenness centrality is the proportion of these paths that go through h. Intuitively, if the betweenness centrality of a hub is significantly larger than the average, this means that the network is centralized around this hub. In the context of the LN, this will mean that a larger proportion of the payments routed in the network will go through a specific hub. Looking through this lens, we will say the network is decentralized if the betweenness centrality of the hubs is similar. To measure this, we look at the standard deviation of the values of the betweenness centrality. The smaller the standard deviation, the better the decentralization of the network.

To be more precise, the cases we have simulated are the following. For each hub in the Network:

(i) **Uniform** The hub chooses k hubs uniformly at random (as prescribed by Maypoles)

(ii) **Preferred Hubs** Let d be the last digit of the hub's ID. The hub chooses k other hubs, where the probability of a hub with a last digit d to be chosen is 10 times greater than that of a hub with the last digit different from d

(iii) 10% **Hub Failure** With probability 90% the hub chooses k hubs uniformly at random, with probability 10% the hub does not add any new connections

(iv) 50% **Hub Failure** With probability 50% the hub chooses k hubs uniformly at random, with probability 50% the hub does not add any new connections

For all of the above we have simulated values of k between 0 and 12. Note that $k = 0$ is just the snapshot of the LN without any changes. The simulation repeated 50 times, and the average of the results was taken.

Figure 5 shows the edge connectivity of the LN as a function of k for the various cases stated above. The LN snapshot has connectivity 8, as is shown in the $k = 0$ column. As k grows, so does the edge connectivity of the network. Notice that, for example, for $k = 12$ in the Uniform case, Theorem 3.3 guarantees edge connectivity of 12. As the base graph has connectivity 8 and we are adding to it a graph with connectivity 12, we could have expected a connectivity of $12 + 8 = 20$. The simulation shows that the average case out performs the theoretical bounds significantly, as the average connectivity is above 27. In Theorem 3.3 we have shown that various properties will hold w.h.p., yet it seems that the average case will be even better than what is stated in the theorem.

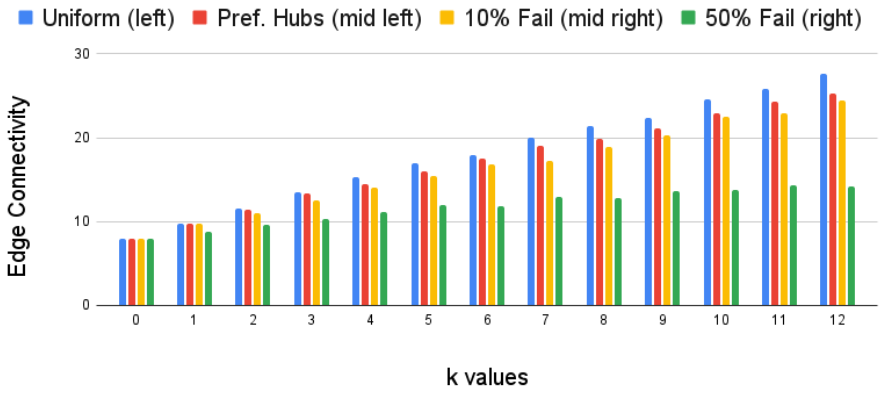

Figure 5: Edge connectivity as a function of the parameter k, for the different variations. The greater the edge connectivity, the better is the topology of the graph.

When comparing Uniform to the other variations in Figure 5, we see that, as expected, nodes that do not cooperate bring the connectivity down. This is particularly significant when the failure probability is 50%. Preferred Hubs on the other hand will not bring connectivity as sharply down, because many edges are added. Other properties studied in Theorem 3.3 exhibit similar improvement dynamics, yet due to the currently small size of the LN, they are less interesting.

With the growth of the network, the diameter and the expansion are expected to become more significant.

Figure 6 shows the standard deviation of the betweenness centrality as a function of k, for the different variation of hub choices. The smaller the standard deviation, the better. A small standard deviation points to a uniform distribution of paths between the hubs. As k grows, we see that the standard deviation becomes smaller. The fact that the routing paths are distributed more evenly between hubs, means that no hub controls a disproportional part of the flow in the network.

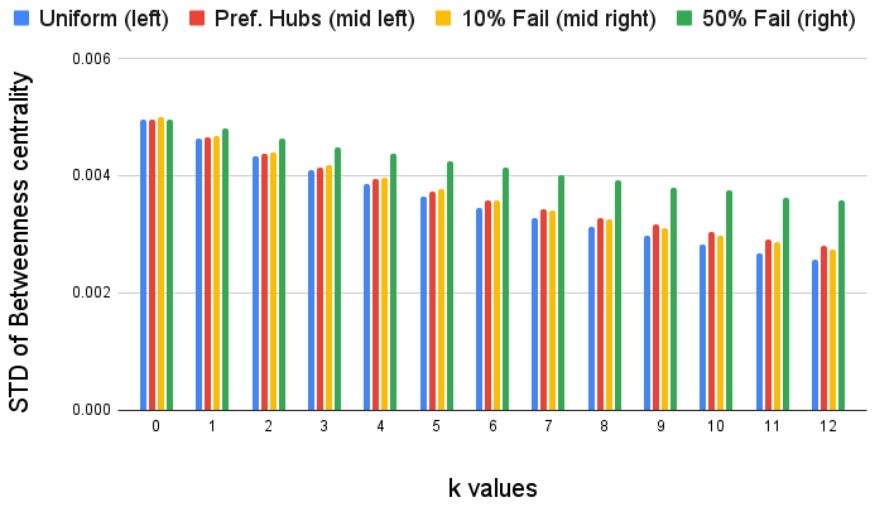

Figure 6: The standard deviation of the betweenness centrality values of hubs in the network as a function of k. The smaller the standard deviation, the better, as it shows the network is more decentralized.

Similarly to the previous figure, the Uniform choice gives the best results, Preferred Hubs and 10% Hub Failure have only slightly larger standard deviations. 50% Hub Failure does not perform as well as the other cases. The hubs that do not cooperate stay behind and do not take part in many paths in the graph, which keeps the standard deviation rather high. For example, 50% Hub Failure for $k = 12$ performs similarly to Uniform with k as small as 5. This points to the fact that hubs that do cooperate get more routes going through them, and this could incentives hubs to join Maypoles.

4 Spokes in Maypoles

In Maypoles, spoke open two channels instead of one. Even if it benefits the network, we cannot expect the spokes to do this without proper economic in-

centive. In this section, we show that by opening channels as prescribed in Maypoles, the spokes save on the costs of continuously transacting over the LN. The savings are due to the dynamics of fees in Bitcoin.

A basic demand from any type of payment system is that it will be always available, that is, at any moment in time the user can send funds that they own and receive funds from other users. Looking at the LN as a mean of payment for spokes, we assume that a spoke in LN needs to always be able to send or receive transactions

In this section we examine the cost of having a channel that is always balanced, that is, always available for sending or receiving transactions. We call such a channel a high availability channel (HAC). We show that following Maypoles allows spokes to have an HAC for a lower cost than trying to have an HAC without Maypoles. The reason that Maypoles is cheaper is that a large part of the cost of an HAC is the on-chain rebalancing fees. When maintaining and using a channel over long periods of time, one must occasionally rebalance by moving funds on the blockchain. Maypoles allows paying smaller fees every time we need to perform such a rebalance, without suffering any downtime.

To gain some intuition, we start with an example. Assume that Alice sells T-shirts online and accepts payments over the LN. Any downtime in which the channel cannot receive payments can result in loss of revenue. Thus, Alice wants a HAC. If Alice has a single channel, when her channel gets depleted, that is, she cannot receive payments, she needs to rebalance. As she has a single channel, any rebalancing option includes an on-chain transaction, and so the main cost will be the transaction fee.

One option is sending the rebalancing transaction with a fee that guarantees it will go through as quickly as possible. This will be expensive, as the waiting time for the transaction to be included strongly depends on the fee. Another option is trying to rebalance in advance, yet this is not cheap either, as it entails locking in the LN large sums (either Alice's or the hub's) and often opening parallel channels to the same hub. Both of these are expensive and in some cases hubs will not allow[5] the spoke to do so.

On the other hand, if Alice's node is a spoke in Maypoles she has two channels, main and secondary. The main channel is the one that she mostly uses to receive payments, and so it is an HAC. Alice's secondary channel is used to rebalance the main channel, that is, when the main channel is depleted Alice uses the secondary channel to rebalance it (see Figure 7 for an illustration). As this rebalance happens on the LN, it is immediate, and the fees are negligible. When the secondary channel needs to be rebalanced, this entails an on-chain transaction. When rebalancing the secondary channel, Alice can wait for a long time for the rebalancing transaction to go through, as her main channel is still open. By allowing longer waiting times when rebalancing, Alice pays a smaller fee. Because of the difference in the rebalancing fees, the cost of maintaining an HAC over Maypoles is smaller than the cost of a single channel as an HAC.

In the next sections we define a general setting for channel costs, prove the

[5]Parallel edges are not even implemented in some LN software

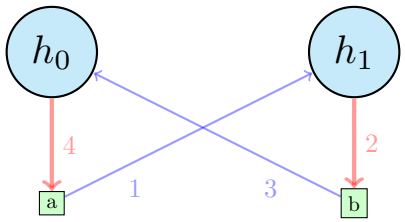

Figure 7: Spoke a is rebalancing the main channel it has with h_0 using the cycle $a \to h_1 \to b \to h_0 \to a$. This also benefits spoke b, as it rebalances its main channel with h_1

benefits for the spokes, and finally give some numeric examples.

4.1 The Cost of a Lightning Channel

The cost of maintaining an HAC has two main parts. The first part is the fees paid each time the channel needs to be rebalanced. The second part is the opportunity loss of the bitcoin locked in the channel. On the one hand, if there is a large amount of bitcoin locked in the channel, we will rarely need to rebalance it, but the opportunity loss is significant. On the other hand, if the amount locked in the channel is a small, so is the opportunity loss, but we need to rebalance the channel often. As the amount of bitcoin locked in the channel can be chosen by the nodes opening it, they can optimize the amount to get a minimal channel cost. As for the rebalancing fee, the smaller it is, the smaller the cost of the channel.

Formally, denote by B the cost of rebalancing a LN channel once, and by $F(B)$ the minimum possible cost of an HAC as a function of B. We assume that the transaction flow rate and the interest rate, that is the rate of opportunity loss, are constant. A natural assumption is that $F(B)$ decreases as B decreases. We use this for the benefit of spokes in Maypoles.

In the LN, and similar solutions, there are two main ways to rebalance the channel. The first one is moving funds from a different channel owned by the same node (see [19] for further details), the second one is performing an on-chain transaction. The LN fees are negligible in comparison to the on-chain fees, and so we may focus on the cost arising from the need to move funds on-chain.

If a node mostly pays on the LN, it will need to move funds from the blockchain to the network regularly, as the funds in all of its channels will run out. Similarly, if a node mostly gets paid, it will need to move funds back to the blockchain to allow for incoming transactions. It is rare that a node sends and receives at exactly the same rate, yet even in this case from time to time the node's funds will be depleted and there will be need for an on-chain transaction to rebalance. See [13] for further details.

Rebalancing through the blockchain entails transaction fees. As there is an ongoing auction for space on the blockchain, a high fee will get the transaction

Figure 8: A moving weekly average of the ratio between the estimated fee for a transaction to appear in 6 blocks and the estimated fee for a transaction to appear in the next block

into the blockchain quickly, while a transaction with a lower fee might wait for hours or days. If a node is not in hurry, it can take advantage of this fact, and send transactions with a lower fee (see [17] for further details). In Figure 8 we show the ratio between the estimated fee for a transaction to be included in 6 blocks versus the next block, between April 2021 and April 2022 as presented in [7]. Waiting 6 blocks cuts the transaction fee by half in most cases. We believe that waiting even longer would allow for even lower fees.

4.2 Benefits of Maypoles for a Spoke

To show that there is a financial motivation for spokes that wish to have an HAC, we need to show that by choosing correctly the sizes of the main and secondary channels in Maypoles, the cost of having an HAC with Maypoles is cheaper in expectation than the cost of a single channel HAC. With a single channel, if one wishes it to have high availability, one needs to either open a new channel before the old one is depleted, thus forgoing interest for large sums in the LN, or to pay high transaction fees. Maypoles avoids this problem and saves on rebalancing costs.

A spoke in Maypoles will hold in its main channel an amount which allows for a few days worth of transactions. The secondary channel would be of size optimized for a minimal cost. The spoke can make sure that the main channel always has liquidity, by moving liquidity from the secondary channel. The move of liquidity from the secondary channel to the main channel is done through a short cycle in the LN, as shown in Figure 7, and so has a negligible cost.

The spoke can wait a long time before rebalancing the secondary channel, as it is only used to rebalance the main channel. Thus, even if this is done by an on-chain transaction, this can be done at a significantly lower cost.

In the following theorem, we show that a spoke's HAC will be cheaper in Maypoles than in the single channel case, if the spoke chooses the sizes of the main and secondary channels correctly. The savings are due to the difference

in the on-chain transaction fees. We define an *immediate* transaction to be a transaction offering a fee high enough for it to be in the next block, and a *delayed* transaction to be a transaction offering a fee high enough for it to appear in the next 6 blocks. The choice of 6 blocks is rather arbitrary and can be a few days worth of waiting, as shown in examples in Section 4.3.

Theorem 4.1. *Let $F(B)$ be the channel cost function, let B_i be the cost of immediate rebalancing and let B_d be the cost of delayed rebalancing. For $B_i > B_d$, there is a choice of sizes for the channels in Maypoles that make an HAC cheaper in Maypoles than a single channel HAC.*

Proof. Let $F_i = F(B_i)$ be the cost of a single channel HAC, and let us compare it to the cost of the Maypoles channels. In Maypoles, we can guarantee that the cost of the secondary channel is $F_d = F(B_d)$. As for the main channel, assume that it is has K bitcoin in it. The size of the main channel, that is, the amount locked inside it, is an upper bound on its cost, as it is never rebalanced on-chain. Thus, the cost of the two Maypoles channels is at most $F_d + K$.

Choose $K < F_i - F_d$, and as $F(B)$ is monotone in B, we know that we can choose $K > 0$. Then we get that

$$F_i > F_d + K$$

that is, the cost of the two Maypoles channels is smaller than the cost of a single regular channel. □

An important observation is that the benefits for a spoke do not depend on the hubs' cooperation. Although rebalancing the spoke's main channel is easier and cheaper if there is a short cycle for the spoke to use, the spoke can make do by finding a path between the two hubs it is connected to. Another solution for the spoke is opening both channels to the same hub, yet this forgoes any privacy benefit and not all hubs allow it.

We finish this section by giving some numeric examples based on the cost functions derived in [13], to show that this indeed works with real world numbers.

4.3 Numeric Examples

Assume that nodes n_0 and n_1 open a channel where funds flow at rate λ_0 from n_0 to n_1 and at rate λ_1 in the opposite direction, and assume that r is the market interest rate. As before, B is the cost of rebalancing the channel once. Consider the cost function whose asymptotic around $r = 0$ are given in [13].

Theorem 4.2 ([13])**.** *In the limit of r near zero, the first order approximation of the minimum cost of a channel is*

(i) If $\lambda_0 > \lambda_1$ then

$$2\left(\frac{B(\lambda_0 - \lambda_1)}{r}\right)^{1/2}$$

(ii) If $\lambda_0 = \lambda_1 = \lambda$ then

$$3\left(\frac{2B\lambda}{r}\right)^{1/3}$$

Example 1 Assume that a spoke spends $\lambda_0 = 10,000$\$ annually over the LN and does not receive funds over the LN, that is $\lambda_1 = 0$. Furthermore, assume that the fee for rebalancing the channel in the next block is $B_i = 1$\$ and the fee for a delayed rebalancing (say, going into a block in the next 24 hours) is $B_d = 0.25$\$ and let the interest rate be $r = 0.05$. Then the expected cost of a single channel is approximately $2\left(\frac{1\cdot 10000}{0.05}\right)^{1/2}\$ \approx 894\$$.

As for the two channel case, assume that in the main channel we deposit 50\$, the amount expected to be used in two days, and in the secondary channel the optimal amount for the minimal cost indicated by Theorem 4.2. Then, in total, the cost is expected to be at most $50\$ + 2\left(\frac{0.25\cdot 10000}{0.05}\right)^{1/2}\$ \approx 274\$$.

From this, we see that the user is expected to save ≈ 620\$ each year, which is almost 70%.

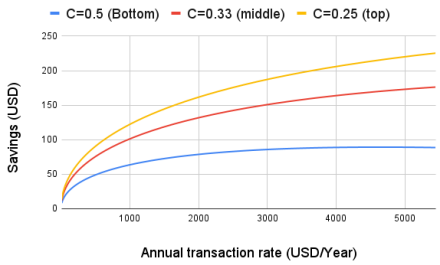

Figure 9: In this graph, we compare the savings in Maypoles for different transaction rates λ, assuming various ratios C between immediate and delayed refunding of a channel.

Example 2 In Figure 9 we examine the amount saved by opening two Maypoles channels instead of one regular channel for various transaction rates and ratios between transaction fees. Define $C = \frac{B_d}{B_i}$, that is, the ratio between a delayed transaction fee and an immediate transaction fee. In this example, we assume that the main channel has a week's worth of funds locked in it. We see that the savings grow as the transaction rate grows and as the ratio C gets smaller.

4.4 Previous Works

This work builds upon previous research on network theory, random graphs, and the economy of the LN. This is the first protocol that offers to use the economic

incentives in the LN to improve its topology. We give a short overview of results used in this paper, directly or indirectly.

The starting point of this paper is the fact that the LN has a hub and spoke topology, and that it is not expected to change. This has been shown in several studies, both empirical (e.g., [10], [27], [20], [21], [15], [31], [32]) and theoretical (e.g., [3], [4], [26]).

It is particularly important to note the studies that point to weaknesses of the network due to its tendency towards the hub and spoke model. See for example [22], [25], [28] and to some extent also [16]. Although some of this works offer local remedies for specific attacks, the hub and spoke topology is fragile by nature, and so in Maypoles we aim to change the topology itself and thus resolve many of the potential problems shown in these papers.

The health of networks is a well studied subject, and there are many works from which we can draw properties we want our network to exhibit. These works can be both theoretical (e.g., [6], [23], [5]), and as practical as network manuals (e.g., [9]). As the LN is decentralized and arises from economic needs, we cannot expect the healthy topologies described in the aforementioned works to appear on their own. To guarantee a good network topology, we need to push the network towards it.

To prove the compatibility with node's economic incentives, we use the results in [17] that show the correlation between waiting times and transaction fees. When going into concrete examples, we use [13] where the authors study closely the cost of lightning channels, together with other questions. This work has only considered the LN on the scope of a channel, we use it to examine how local economic incentives can be used to better the network in general.

5 Conclusion and Discussion

In this work, we have introduced and studied the Maypoles protocol, which improves payment channel networks that have a hub and spoke topology, such as the Lightning Network. We have shown that Maypoles increases the network stability, privacy, and resilience to attacks. Going over several parameters that measure the health of a network, we have shown that both in theory and in practice, Maypoles improves them significantly.

We have also shown that by correctly choosing the channel sizes, the spokes can save on costs of maintaining a channel that is always available for sending and receiving transactions The savings are due to the correlation between transaction fee over Bitcoin and the urgency of the transaction. This can motivate spokes to participate in the Maypoles protocol. By taking into account the economic incentives, we made sure to create a win-win situation between individual nodes and the network as a whole.

The framework created here can be used to examine other variation of decentralized changes to the topology. In Section 3.3 we have discussed a few variations on the Maypoles protocol and simulated their performance. In the future, it could be interesting to examine other variations tailored to the specific

challenges of different networks.

It could also be interesting to examine the addition of channels to other topologies, such as scale-free graphs (see, e.g, [1]). In such networks, there won't be a clear-cut between "large" nodes, like hubs, and "small" ones, like spokes. This might prove to be more useful for the study of some payment channel networks, especially as they grow.

Another interesting direction to look into is encouraging nodes to open more than one channel. In Maypoles, we give economic incentive for opening two channels. It would be better if the number of channels would be even greater. Opening more than two channels as described in Maypoles will not offer significant savings. It could be that there is another scheme which can motivate spokes to have many channels.

The LN is a very young project with many changes and upgrades happening. For example, Multi-Path Payment is a solution that allows splitting a single payment into several smaller payments and sending the smaller payments along different routes. The ability to do so can motivate nodes to open several channels for better privacy and liquidity management. Studying how to better utilize this and other developments for the health of the network would be of great importance to the development of the LN.

References

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[2] Removed for anonymity. Maypoles simulation. `https://https://github.com/`, 2022.

[3] Silvia Bartolucci, Fabio Caccioli, and Pierpaolo Vivo. A percolation model for the emergence of the bitcoin lightning network. *Scientific Reports*, 10(1):1–14, 2020.

[4] Ferenc Béres, Istvan Andras Seres, and András A Benczúr. A cryptoeconomic traffic analysis of bitcoins lightning network. *arXiv preprint arXiv:1911.09432*, 2019.

[5] Laxmi N Bhuyan and Dharma P Agrawal. A general class of processor interconnection strategies. *ACM SIGARCH Computer Architecture News*, 10(3):90–98, 1982.

[6] Laxmi N. Bhuyan and Dharma P. Agrawal. Generalized hypercube and hyperbus structures for a computer network. *IEEE Computer Architecture Letters*, 33(04):323–333, 1984.

[7] Bryan Aulds Colin Aulds. Bitcoin fee estimator, 2022.

[8] Christian Decker. Lightning network research - topology datasets. `https://github.com/lnresearch/topology`. Accessed: 2020-10-01.

[9] Cisco Validated Design. Campus network for high availability design guide, 2008.

[10] Guillaume Felley, Arthur Gervais, and Roger Wattenhofer. Towards usable off-chain payments. 2018.

[11] Trevor I. Fenner and Alan M. Frieze. On the connectivity of random m-orientable graphs and digraphs. *Combinatorica*, 2(4):347–359, 1982.

[12] Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.

[13] Paolo Guasoni, Gur Huberman, and Clara Shikhelman. Lightning network economics: Channels. *to appear in Management Science*, 2023.

[14] Lewis Gudgeon, Pedro Moreno-Sanchez, Stefanie Roos, Patrick McCorry, and Arthur Gervais. Sok: Layer-two blockchain protocols. Cryptology ePrint Archive, Report 2019/360, 2019. `https://ia.cr/2019/360`.

[15] Yuwei Guo, Jinfeng Tong, and Chen Feng. A measurement study of bitcoin lightning network. In *2019 IEEE International Conference on Blockchain (Blockchain)*, pages 202–211. IEEE, 2019.

[16] Jona Harris and Aviv Zohar. Flood & loot: A systemic attack on the lightning network. In *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies*, pages 202–213, 2020.

[17] Gur Huberman, Jacob Leshno, and Ciamac C Moallemi. Monopoly without a monopolist: An economic analysis of the bitcoin payment system. *Bank of Finland Research Discussion Paper*, (27), 2017.

[18] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*, volume 45. John Wiley & Sons, 2011.

[19] Rami Khalil and Arthur Gervais. Revive: Rebalancing off-blockchain payment networks. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 439–453, 2017.

[20] Seungjin Lee and Hyoungshick Kim. On the robustness of lightning network in bitcoin. *Pervasive and Mobile Computing*, 61:101108, 2020.

[21] Jian-Hong Lin, Kevin Primicerio, Tiziano Squartini, Christian Decker, and Claudio J Tessone. Lightning network: a second path towards centralisation of the bitcoin economy. *arXiv preprint arXiv:2002.02819*, 2020.

[22] Dmytro Piatkivskyi, Stefan Axelsson, and Mariusz Nowostawski. Digital forensic implications of collusion attacks on the lightning network. In *IFIP International Conference on Digital Forensics*, pages 133–147. Springer, 2017.

[23] Samuel Pierre and Gisele Legault. A genetic algorithm for designing distributed computer network topologies. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 28(2):249–258, 1998.

[24] Joseph Poon and Thaddeus Dryja. The bitcoin lightning network: Scalable off-chain instant payments, 2015.

[25] Elias Rohrer, Julian Malliaris, and Florian Tschorsch. Discharged payment channels: Quantifying the lightning network's resilience to topology-based attacks. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 347–356. IEEE, 2019.

[26] Sascha Schmid. Balanced routing in micropayment channel networks. 2017.

[27] István András Seres, László Gulyás, Dániel A Nagy, and Péter Burcsi. Topological analysis of bitcoin's lightning network. In *Mathematical Research for Blockchain Economy*, pages 1–12. Springer, 2020.

[28] Sergei Tikhomirov, Pedro Moreno-Sanchez, and Matteo Maffei. A quantitative analysis of security, anonymity and scalability for the lightning network. *IACR Cryptol. ePrint Arch.*, 2020:303, 2020.

[29] Florian Tramèr, Dan Boneh, and Kenny Paterson. Remote {Side-Channel} attacks on anonymous transactions. In *29th USENIX security symposium (USENIX security 20)*, pages 2739–2756, 2020.

[30] Aaron Van Wirdum. *Understanding The Lightning Network*, 2016. `https://bitcoinmagazine.com/articles/understanding-the-lightning-network-part-building-a-bidirectional-payment-channel-14647`

[31] Jie Wu and Suhan Jiang. Local pooling of connected supernodes in lightning networks for blockchains. In *2020 IEEE International Conference on Blockchain (Blockchain)*, pages 421–427. IEEE, 2020.

[32] Philipp Zabka, Klaus-Tycho Foerster, Christian Decker, and Stefan Schmid. Short paper: A centrality analysis of the lightning network. 2022.

A The Lightning Network

A channel in the LN is opened between two parties. Both parties lock a sum in the channel by signing an on-chain Bitcoin transaction. Once the channel is open, they can transact between themselves by changing the ownership over parts of the locked funds.

For example, assume that Alice regularly buys coffee from Bob. They open a channel where Alice locks 5 bitcoin that she is planning to pay Bob with. Bob is not locking any funds, as he does not expect to pay Alice.

The first time Alice buys coffee, she wants to transfer 1 bitcoin to Bob, and so they change the state of the channel to Alice having 4 bitcoin and Bob having

1 (see Figure 10). They can continue transacting if the channel's funds allow it, and Bob can also send bitcoin to Alice if needed.

Figure 10: An example of Alice paying Bob 1 bitcoin over their channel

If there is also a channel between Bob and Carol, Alice can transact with Carol without opening a new channel to her. She can pay Carol through Bob, if Bob agrees to cooperate. Bob can charge a fee for the service he is providing to Alice and Carol. To pay Carol 1 bitcoin, Alice sends 1 bitcoin to Bob, who then sends 1 Bitcoin to Carol. See Figure 11 for this example.

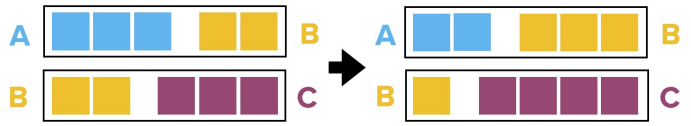

Figure 11: An example of Alice sending Carol 1 bitcoin through Bob

The last example we want to consider is when Alice has channels both to Bob and to Carol, and Bob and Carol have their own channel. Assume Alice and Bob's channel is depleted, that is, Bob owns all the bitcoin, and assume that Alice still has funds in her channel with Carol. Then Alice can refund her channel with Bob by moving some funds from the Alice and Carol channel to the Alice and Bob channel. Bob and Carol must agree to this, and the channel their channel must have sufficient funds.

For example, if Alice wants to send 1 bitcoin from her channel with Carol to her channel with Bob, she will do this by using Carol and Bob's channel. To be more precise, Alice sends herself 1 Bitcoin through Carol and Bob. In the channel of Alice and Carol, Alice sends 1 Bitcoin to Carol. In the channel of Carol and Bob, Carol sends 1 bitcoin to Bob. In the Channel of Bob and Alice, Bob sends 1 bitcoin to Alice. Each one has the same amount in the beginning and the end, only the distribution between the channels changed. See Figure 12 for an illustration.

Note that this is cryptographically guaranteed to ensure that no harm can come to any party. If Alice pays Bob, she cannot later claim that the transaction did not happen, and if Bob and Carol agree to facilitate the transaction, they cannot disappear with the funds and not commit their part of the deal. The full details can be found in [24] and in several blog posts, such as [30].

B Proof of The Main Theorem

A key part of the proof of Theorem 3.3 is the observation that the Maypoles protocol gives raise to a k-out graph, as introduced in [11].

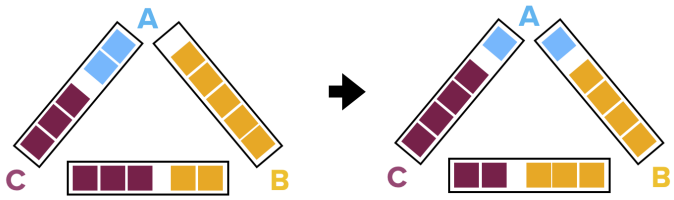

Figure 12: An example of Alice rebalancing her channel with Bob, by using the funds she has in the channel with Carol

Definition B.1 (k-out graph)**.** *Starting from an empty graph $G = (V, \emptyset)$ we do the following steps*

(i) For each $v \in V(G)$ create a set $A(v)$ by choosing k nodes from $V(G)$ uniformly, independently, and allowing repetitions

(ii) For each $v \in V(G)$ add an edge between v and each node in $A(v)$

(iii) Remove double edges and self-loops

The resulting graph is called a k-out graph.

The proof of the main theorem builds upon the observation that $\mathcal{L}_k$ is a k-out graph. To prove the claimed graph properties, we need the following two lemmas. The first lemma is in the heart of the theorem, and shows that a k-out graph has good expansion. The second lemma shows that graphs with good expansion have a low diameter. We start by stating and proving the lemmas, and then we proceed to the proof of the theorem.

Lemma B.2. *Let $G = (V, E)$ be a random k-out graph, for some constant k. Then for every $S \subset V$, such that $|S| \leq |V|/2$, we have that $\partial(S) \geq (1+o(1))k|S|$ w.h.p.*

Proof. We split the proof into two cases. We first focus on the more complicated case where $|S| > n^{1/4}$. After proving that the lemma holds for such S, we finish the proof by showing the simpler case where $|S| \leq n^{1/4}$.

Let $|S| = s$ and $|V| = n$. Define $\overline{\partial}(S)$ to be the random variable counting the number of edges between S and $V \setminus S$ prior to the removal of double edges in the k-out processes, that is, prior to step (iii) of the processes. We will show that w.h.p. $\partial(S) = (1 + o(1))\overline{\partial}(S)$ and $\overline{\partial}(S) \geq (1 + o(1))ks$.

To show that w.h.p. $\overline{\partial}(S) \geq (1 + o(1))ks$, we calculate the expectation of $\overline{\partial}(S)$ and then show that $\overline{\partial}(S)$ is concentration around its expectation. Edges in $\overline{\partial}(S)$ appear if either a node in S chooses a node in $V \setminus S$ or if a node in $V \setminus S$ chooses a node in S, call the former type 1 and the latter type 2.

For each $v \in S$ the expected number of nodes it chooses outside S is

$$k \cdot \frac{|V \setminus S|}{|V|}$$

Thus the expected number of edges of type 1 is

$$|S| \cdot k \cdot \frac{|V \setminus S|}{|V|} = k\frac{s(n-s)}{n}.$$

Similarly, the expected number of edges of type 2 is

$$|V \setminus S| \cdot k \cdot \frac{|S|}{|V|} = k\frac{s(n-s)}{n}.$$

Thus, the expected number of edges with one node in S and one node in $V \setminus S$ is

$$\mathbb{E}[\overline{\partial}(S)] = 2k\frac{s(n-s)}{n}$$

To show that $\overline{\partial}(S)$ is concentrated around its expectation we use Chebyshev's inequality

Theorem (Chebyshev's inequality) *For any random variable X and positive a*

$$\mathbb{P}(|X - \mathbb{E}[X]| > a) \leq \frac{Var[X]}{a^2}$$

The variance of the random variable $\overline{\partial}(S)$ can be calculated as follows. For $v \in S$ let X_v be the number of nodes in $V \setminus S$ chosen by v, and for $u \in V \setminus S$ let Y_u be the number of nodes in S chosen by u. Note that the above variables are all independent of each other, and that $\overline{\partial}(S) = \sum_{v \in S} X_v + \sum_{u \in V \setminus S} Y_u$. Using these two facts we get that

$$Var[\overline{\partial}(S)] = \sum_{v \in S} Var[X_v] + \sum_{u \in V \setminus S} Var[Y_u].$$

As X_v behaves as the Binomial random variable $Bin(k, \frac{n-s}{n})$, and Y_u behaves as $Bin(k, \frac{s}{n})$, we get

$$\begin{aligned} Var[\overline{\partial}(S)] &= \sum_{v \in S} k\frac{n-s}{n}(1 - \frac{n-s}{n}) + \sum_{u \in V \setminus S} k\frac{s}{n}(1 - \frac{s}{n}) \\ &= sk\frac{n-s}{n} \cdot \frac{s}{n} + (n-s)k\frac{s}{n} \cdot \frac{n-s}{n} \\ &= n \cdot k\frac{s}{n}\frac{(n-s)}{n} = k\frac{s(n-s)}{n}. \end{aligned}$$

Notice that $Var[\overline{\partial}(S)] = \frac{1}{2}\mathbb{E}[\overline{\partial}(S)]$, and so by choosing $a = \mathbb{E}(\overline{\partial}(S))^{2/3}$ and plugging it into Chebyshev's inequality we get

$$\mathbb{P}[|\overline{\partial}(S) - \mathbb{E}[\overline{\partial}(S)]| > \mathbb{E}[\overline{\partial}(S)]^{2/3}] \leq \frac{Var[\overline{\partial}(S)]}{\mathbb{E}[\overline{\partial}(S)]^{4/3}} = \frac{1}{2\mathbb{E}[\overline{\partial}(S)]^{1/3}}$$

It is left to show that $\frac{1}{2\mathbb{E}[\overline{\partial}(S)]^{1/3}} \underset{n\to\infty}{\longrightarrow} 0$. As $\frac{n}{2} \geq s \geq n^{1/4}$, and as $f(s) = s(n-s)$ is a monotone increasing function in this range, we have that

$$\begin{aligned}(2\mathbb{E}[\overline{\partial}(S)])^{-1/3} =& (4k\frac{s(n-s)}{n})^{-1/3} \\ \leq& (4k\frac{n^{1/4}\cdot(n-n^{1/4})}{n})^{-1/3} \leq n^{-1/12} \to 0.\end{aligned}$$

Thus we have that w.h.p.

$$\begin{aligned}\overline{\partial}(S) &= (1+o(1))\mathbb{E}[\overline{\partial}(S)] = (1+o(1))2k\frac{s(n-s)}{n} \\ &\geq (1+o(1))2k\frac{s(n-n/2)}{n} = (1+o(1))ks \qquad (1)\end{aligned}$$

where inequality (1) holds as $s \leq \frac{n}{2}$.

As for the case where $s \leq n^{1/4}$, the probability that a node in S chooses a node in $V \setminus S$ is

$$\frac{|V\setminus S|}{|V|} = \frac{n-s}{n} \geq \frac{n-n^{1/4}}{n} = 1 - \frac{1}{n^{3/4}}.$$

The probability that all the nodes in S choose nodes in $V \setminus S$ is at least

$$(1-\frac{1}{n^{3/4}})^{ks} \geq (1-\frac{1}{n^{3/4}})^{kn^{1/4}} \underset{n\to\infty}{\longrightarrow} 1.$$

From the above, if $s \leq n^{1/4}$ we have that w.h.p. all the edges that nodes in S chose are to $V \setminus S$. This means that w.h.p. $\overline{\partial}(S) \geq ks$.

It is left to show that $\partial(S) = (1+o(1))\overline{\partial}(S)$. The difference between $\overline{\partial}(S)$ and $\partial(S)$ is the number of double edges between S and $V \setminus S$. Denote the random variable counting these double edges by $D(S)$, and note that

$$\partial(S) \geq \overline{\partial}(S) - D(S). \qquad (2)$$

Let $v \in S$ and $u \in V \setminus S$. The probability of more than one edge between them is at least the probability of exactly 2 edges. The nodes v and u make together $2k$ choices. The probability that u chooses v or vice versa is $\frac{1}{n}$. Thus, the number of edges between them follows the binomial distribution $Bin(2k, \frac{1}{n})$ and so

$$\mathbb{P}(\text{exactly 2 edges between } u \text{ and } v) = \binom{2k}{2}\frac{1}{n^2}(1-\frac{1}{n})^{2k-2} \leq \frac{4k^2}{n^2}.$$

From this, we get that the expected number of double edges is at most

$$\mathbb{E}(D(S)) \leq s(n-s)4\frac{k^2}{n^2} = \frac{2k}{n}\mathbb{E}[\overline{\partial}(S)].$$

By Markov's inequality we have that

$$\mathbb{P}\left(D(S) > n^{1/2}\mathbb{E}(D(S))\right) \leq \frac{\mathbb{E}(D(S))}{n^{1/2}\mathbb{E}(D(S))} = \frac{1}{n^{1/2}} \to 0$$

and so w.h.p. we have that $D(S) \leq n^{1/2}\mathbb{E}[D(S)] \leq n^{1/2}\frac{2k}{n}\mathbb{E}[\overline{\partial}(S)] = o(\mathbb{E}[\overline{\partial}(S)])$.

Plugging the above into (2), we get that w.h.p.

$$\partial(S) \geq (1 + o(1))\overline{\partial}(S) \geq (1 + o(1))ks$$

and this completes the proof.

□

Lemma B.3. *Let G be a graph on n nodes. If G is a k-expander, then the diameter of G is at most* $\log n/(k+1)$

Proof. For any $v \in V(g)$, let $S_i(v)$ be the set of nodes of distance at most i. $S_0(v) = |\{v\}| = 1$. As G is a k-expander, the number of nodes in $S_1(v)$, including v itself, is at least $S_0(v) + k \cdot S_0(v) = k + 1$. If $S_{m-1}(v) < \frac{n}{2}$ then the number of nodes in $S_m(v)$ is at least $S_{m-1} + k \cdot S_{m-1} \geq (k+1)^m$. Choosing $m_0 = \frac{1}{2}\log n/\log(k+1)$, we get that there are at least $n/2$ nodes of distance at most m_0 from v.

Let u and v be a pair of nodes in G. From the above, there are at least $n/2$ nodes of distance at most m_0 from u and the same holds for v. Thus, there must be a node in both $S_{m_0}(v)$ and $S_{m_0}(u)$, and so there is a path of length at most $2m_0 = \log n/\log(k+1)$ from u to v. □

We are now ready to prove Theorem 3.3.

Proof of Theorem 3.3. The key observation we need is that $\mathcal{L}_k$ is equivalent to a k-out graph. Indeed, as a hub instructs its spokes to connect to k different hubs that are chosen uniformly at random, it creates k random edges in $\mathcal{L}_k$. This is exactly step (ii) of the k-out creation processes. If two hubs choose each other, or a hub chooses another hub twice, we still think about the connection between them as a single edge in $\mathcal{L}_k$, and this is step (iii) of the k-out processes. This shows that the resulting graph $\mathcal{L}_k$ is a k-out graph.

Items ((i)) and ((ii)) are obtained by the above observation and the following result of Frieze.

Theorem[[12], Theorem 17.2] Let $k \geq 2$ be a fixed integer. Then a k-out graph has w.h.p. edge connectivity and node connectivity k.

Item ((iii)) is a direct consequence of the above observation and of Lemma B.2 that shows that k-out graphs are$(1 + o(1))k$ expanders. Item ((iv)) is obtained by the observation, Lemma B.2, and Lemma B.3 that shows that expanders have the needed diameter. □