Resilient Onion Messaging in the Lightning Network

Amin Bashiri, Majid Khabbazian

***Abstract*—Onion messages (OMs) in the Lightning Network (LN) enable private communication between nodes through onion routing. While they support important functionalities such as static invoices and asynchronous payments, they can also be exploited for spam. To counter this, the Basis of Lightning Technology Specifications (BOLTs) recommend implementing rate limiting on OM forwarding. Limiting the rate, however, opens the door for an adversary to degrade the OM service by flooding the network—anonymously, thanks to onion routing.**

In this paper, we analyze and quantify the impact of DoS attacks on the OM service. Following the approach suggested in the literature, we study a policy in which per-peer forwarding limits and forwarding-node selection probabilities are proportional to publicly observable channel-capacity weights. Our analysis shows that OMs are most resilient against DoS attacks when the honest nodes' public-capacity weights are uniformly distributed. However, the distribution of these weights in LN is highly non-uniform, as demonstrated in prior empirical studies and confirmed by our analysis of three network snapshots spanning 2022–2026. To improve resilience under such skewed conditions, we propose restricting forwarding-node selection to a carefully selected subset of nodes.

For the evaluated setting $l = L = 3$, at a common adversarial public-capacity weight of approximately 14.30 BTC—equivalent to about USD 1 million at the reference BTC/USD rate used in our evaluation—the top-target per-connection adversary matches the strong optimal aggregate-budget adversary in each snapshot. Their OM failure probabilities range from 40.9% to 56.7% when the full network is available for forwarding-node selection. Restricting selection to the optimized subsets reduces these probabilities to 3.6%–3.9%, with corresponding analytical upper bounds of 6.2%–6.9%.

***Index Terms*—Blockchain, Bitcoin, Lightning Network, Onion Messages, Denial of Service**

I. Introduction

The Lightning Network (LN) [1] is a payment channel network [2], [3] launched to address Bitcoin's scalability challenges. In LN, each channel is formed by two users who lock funds in a funding transaction on the Bitcoin blockchain. Through these channels, users can send payments to each other without the need to record a transaction to the Bitcoin blockchain.

The payment process on LN currently operates with unidirectional communication prior to executing each payment. In this process, the payee generates an invoice containing all the details required to complete the payment and sends it to the payer. This invoice is often communicated off-network, for example, through scanning a QR code. Upon receiving the invoice, the payer can then execute the payment using the information provided.

In the above process, the only communication before payment execution is from the payee to the payer through the invoice. Many useful applications, however, require short messages in both directions between payer and payee, either before or after a payment. One prominent example is so-called static invoices: a static invoice is a reusable, long-lived payment descriptor (often encoded as a QR code or URI) that a payee can display once and that payers can scan multiple times; each scan causes the payer's wallet to contact the payee over onion messages and request a fresh, per-payment invoice with the desired amount and metadata [4], [5].

Refunds follow a similar pattern. From the protocol's point of view, a refund is just a second payment in the opposite direction, but initiating it requires the new payer (the original payee) to obtain refund destination information (e.g., an invoice or blinded path) from the new payee (the original payer). In practice this reverse-direction message is often sent over auxiliary channels such as email, web support forms, or proprietary app messaging. Onion messages provide a protocol-native, onion-routed channel for these bidirectional interactions, without requiring the payer to operate an HTTP endpoint or reveal additional contact information. As LN evolves into a comprehensive financial platform, facilitating these use cases is becoming more essential.

To enable bidirectional communication, one proposed solution is the LNURL (Lightning Network Uniform Resource Locator) [6]. LNURL integrates HTTP servers with LN nodes to facilitate two-way communication. This integration, however, necessitates that users run an HTTP server alongside their LN node, which increases the node's resource requirements, adds operational complexity, enlarges the attack surface, and potentially compromises user privacy and anonymity.

A more recent alternative is onion messaging [7], which utilizes onion routing–similar to the approach used by Tor–to enable secure bidirectional communication. Onion routing [8] provides strong anonymity and privacy for users. However, it also allows adversaries to misuse the OM service, such as sending spam, without fear of detection. To counter this, the Basis of Lightning Technology Specifications (BOLTs) recommend implementing rate limiting on OM forwarding. However, limiting the rate introduces a potential vulnerability: an adversary can degrade the OM service by flooding the network with excessive traffic—anonymously, thanks to onion routing. This is conceptually similar to channel jamming attacks in LN, where adversaries disrupt payment routing by saturating channels with minimal cost, as demonstrated by

Amin Bashiri and Majid Khabbazian are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada (e-mail: abashiri@ualberta.ca; mkhabbazian@ualberta.ca).

Shikhelman and Tikhomirov [9].

An earlier short paper by the same authors [10] introduced the OM flooding threat and studied a baseline mitigation in which forwarding limits and forwarding-node selection probabilities are proportional to node funds, together with hop-distance leashes. It provided a high-level analysis under simplifying assumptions, including a complete communication graph, equal funds across nodes, and a small number of adversarial nodes.

The present paper retains fund-proportional forwarding and routing as the baseline mechanism, but substantially extends the earlier work. We analyze arbitrary non-uniform fund distributions and derive new failure-probability bounds in terms of the adversary's aggregate fund share and the skew of the honest fund distribution. We also characterize an optimal link-saturation strategy and introduce a new system design that restricts forwarding-node selection to a carefully selected subset of nodes. For this design, we derive a new resilience bound and provide a quadratic-time algorithm for finding an optimal forwarding subset. Finally, the stronger budget-based adversarial analysis, the formal privacy analysis, and the evaluation over real LN snapshots are all new to this paper.

Relation to alternative defenses. Other defenses address different aspects of OM flooding. Fixed per-peer rate limits are simple, but assigning the same quota to every peer may permit Sybil amplification unless separate admission constraints are imposed. Hop-distance leashes, including hard hop limits and proof-of-work-based soft leashes [10], reduce the amplification obtained from long paths but do not prevent an adversary from sending many short OMs. A feedback-based per-peer throttling proposal [11] propagates congestion notifications backward and reactively reduces the quota of the peer believed to be responsible. This reactive local mechanism is complementary to our proactive capacity-weighted forwarding and subset selection, and the two approaches could be combined. Because these designs rely on different resources and assumptions, we compare them qualitatively rather than numerically.

II. System Model

We model LN using a graph $G_L = (V, E_L)$, where V represents the set of nodes in the network, and E_L represents the set of public channels. The communication graph is also represented as a graph $G_C = (V, E_C)$, where E_C denotes the communication links between nodes in LN.

Definition 1 (Node Fund)**.** *For purposes of this paper, the fund F_i of node U_i is defined as the aggregate capacity of all public channels incident to U_i, as observed from the public LN graph. Thus, F_i is a public-capacity weight, not the node's private channel balance or instantaneous payment liquidity. It is used to determine OM forwarding quotas and path-selection probabilities, not to estimate the node's ability to forward payments.*

We assume that each node in LN maintains a list of all other nodes in the network, along with the amount of their funds as defined above. This assumption aligns with LN's gossip protocol, where honest nodes verify that announced channels correspond to an existing funding transaction before adding them to their local view of the network. Under these standard validation rules, an adversary cannot increase its apparent funds merely by advertising fake link capacities: any increase in its "fund" value must be backed by actually locking additional coins on-chain. Attacks in which honest nodes are tricked into accepting fictitious channels that are not backed by a valid funding output (and hence fictitious capacities) are ruled out by our model, as they would violate the underlying Bitcoin and LN assumptions on which all routing already relies. Although this capacity is backed by valid on-chain funding outputs, it does not reveal how the corresponding channel balances are divided between their endpoints and should not be interpreted as funds entirely owned or controlled by the node.

Forwarding. Onion-message rate limiting is already supported as a local mechanism for controlling spam: a node may limit OMs that it is asked to forward. Our model does not introduce a new rate-limiting mechanism, but specifies how the per-peer forwarding quota is determined. In particular, rather than applying a uniform quota, a node accepts forwarding requests from peer U_i at a rate of at most αF_i, where α is a system parameter and F_i is the aggregate capacity of U_i's public channels. This value is publicly computable from the LN graph and its corresponding on-chain funding outputs; it does not represent the private or time-varying balance on either side of a channel.

We further assume that nodes do not limit OMs destined for themselves, always prioritize sending their own OMs over forwarding requests from their peers, and do not limit the rate of OMs they send on an outgoing link.

Routing: The Path Selection Algorithm. Fund-weighted path selection is a local sender-side policy. Since an OM sender constructs the path locally, it can assign selection probabilities using the publicly computable channel-capacity values defined above, without requiring coordination among forwarding nodes or the disclosure of private channel balances.

The routing algorithm, which we refer to as the *random-path selection algorithm*, selects a random path of length l by independently choosing $l-1$ forwarding nodes, with replacement. The probability of selecting each node is proportional to its funds. Implementing this policy requires a modification to the sender's relay-selection policy, but no change to the onion-message wire protocol or to the Bitcoin consensus rules.

Adversary. The adversary can create one or more nodes in the network by paying the fees for funding transactions and putting in the minimum channel capacity required and can send OMs using these nodes at any rate it wishes. We also assume that joining LN does not present a major barrier to the adversary other than paying the funding transaction fee, but controlling a large portion of the total funds in the network does. Thus, our mitigation method is *Proof-of-Stake (PoS)-like*: a node's OM forwarding rate and its probability of being chosen on a path are both proportional to its economic stake (channel funds), thus creating many low-fund Sybil nodes does not increase the adversary's aggregate influence, in spirit similar to PoS blockchain protocols where a participant's influence scales with its stake rather than the number of

identities it creates.

Multiple real-world attackers that share the same goal (e.g., a coordinated DoS/DDoS campaign) are modeled as a single adversary whose total funds F_A equal the sum of the funds of all nodes participating in the attack; our bounds depend only on this aggregate F_A, not on how it is partitioned across identities.

DoS attacks. Considering the above model, we explore two DoS attacks. In the first attack, the adversary attempts to flood the network with unnecessary OMs to degrade the OM service. This type of attack seeks to hinder the network's general functionality and its ability to efficiently transmit messages.

In the second attack, the adversary targets a victim node with OMs to prevent other nodes in the network from reaching it. This targeted approach seeks to incapacitate a network participant.

To amplify the overall effectiveness of the attacks (i.e., to increase the likelihood of OM delivery failures), we assume in both scenarios that any adversary-controlled node drops any OM originating from an honest node that passes through it.

Resilience metric. Following standard reliability measures in anonymous communication and overlay routing (e.g., the probability that an end-to-end path remains usable in the presence of failures or adversarial relays) [12], [13] and our earlier work on OM DoS in LN [10] we quantify the resilience of the OM service (also referred to as the network's resilience) by the probability that an OM is successfully delivered between a pair of honest nodes. This probability is evaluated over the randomness of the path selection algorithm and is independent of the specific choice of honest nodes.

Table I lists the symbols used throughout the rest of the paper.

TABLE I: Frequently used symbols.

Symbol	Meaning
F_i	Public-capacity weight of honest node U_i, defined as the aggregate capacity of its public channels.
$F_A,\ F_H$	Aggregate public-capacity weights of adversarial / honest nodes.
f_i	Normalised fund share $F_i/(F_A+F_H)$.
L	Maximum hop count an Onion Message may traverse.
l	Path-length parameter of the random-path algorithm ($l \le L$).
γ	Adversary-fund fraction $F_A/(F_A+F_H)$.
$F_{\max}$	Largest single honest balance $\max_i F_i$.
F_{avg}	Average honest balance F_H/n.
δ	Fund-skew factor $F_{\max}/F_{\text{avg}}$.
k	Number of highest-weight honest nodes used in Theorem 2.
$F_{\max}^{(k)}$	Mean balance of those k richest honest nodes.
$\check{\delta}$	Refined skew factor $F_{\max}^{(k)}/F_{\text{avg}}$.
S	Forwarding subset chosen by Algorithm 1.
$F_{\text{tot}}^{(S)}$	Aggregate funds of nodes in S.
$F_{\max}^{(S)}$	Largest balance inside subset S.
α	Rate-limit constant (messages $\cdot$ s^{-1} per satoshi).
$\mathfrak{b}$	Aggregate traffic budget of the adversary (total rate it can inject across honest links).
$\mathcal{E}_{\text{drp}}$	Event that the randomly selected path contains at least one adversarial node.
$\mathcal{E}_{\text{sat}}$	Event that the randomly selected path traverses an honest link—excluding the first and last hop—saturated by the adversary.
$I_{i,j}$	Indicator that link (U_i, U_j) is saturated by the adversary.

III. Resilience Analysis

In this section, we analyze the resilience of the OM service against the DoS attacks described in the previous section. We start with the first attack, in which the adversary attempts to degrade the availability of the OM service by flooding the network.

A. The Flooding Attack

In this attack scenario, the adversary aims to degrade the performance of the OM service by overwhelming the network with a high volume of OMs. To analyze the resilience of the OM service, we start with the assumption that the communication graph is complete. This means that any two nodes in the network can establish a communication link (e.g., a TCP connection) to exchange OMs without requiring a channel between them to forward a message. In Section IV, we retain this pairwise communication assumption but restrict forwarding-node selection to a carefully selected subset of nodes.

Suppose U_s and U_d are two honest nodes in the network. To send an OM to U_d, U_s selects a path using the random-path selection algorithm, which chooses a random path of length l by independently sampling $l-1$ intermediate nodes. The probability of selecting each node is proportional to its total funds.

We say that the OM fails if the selected path satisfies either of the following two conditions: 1) it includes a node controlled by the adversary, which drops the message; or 2) it includes an honest link that has been saturated by adversarial traffic, causing the message to be dropped due to rate-limiting constraints.

The following theorem provides an upper bound on the probability of such a failure. This bound relies on the assumption that the adversary's total funds are limited, representing only a fraction of the total funds in the network.

For instance, while the adversary may create many nodes within the network, the total funds must be distributed across these nodes, resulting in each adversary-controlled node holding only a fraction of the total adversary funds. This directly reduces the likelihood that any individual adversary-controlled node is selected in the random-path selection algorithm, as the selection probability is proportional to the node's available funds. This design inherently limits the collective influence of adversary-controlled nodes, even if they are numerous.

Similarly, the total damage the adversary can inflict by congesting the network is constrained. This is because the random-path selection algorithm tends to favor links that support higher forwarding rates, and the maximum forwarding capacity of each node is also proportional to its funds. As a result, the adversary's ability to congest many high-capacity links remains fundamentally limited by the adversary's total funds.

Theorem 1. *Let L denote the maximum number of hops an OM can travel, and $l \le L$ denote the length of paths set in the random-path selection algorithm. Let F_A be the total funds of the adversary (possibly representing several colluding attackers), and $F_H = \sum_{i=1}^{n} F_i$ be the total fund of honest*

nodes, where F_i denotes the fund of honest node U_i, and n is the total number of honest nodes in the network.
Then, for every pair of distinct honest nodes U_s and U_d, the probability that an OM message, sent on a random path chosen by the random-path selection algorithm, from U_s to U_d fails is at most

$$1-(1-\gamma)^{l-1}+(L-1)\cdot(l-2)\cdot(1-\gamma)\cdot\gamma\cdot\delta \leq (L-1)\cdot(l-1)\cdot\delta\cdot\gamma$$

where

$$\gamma=\frac{F_A}{F_A+F_H}$$

is the fraction of funds in the entire network that belongs to the adversary, and

$$\delta=\frac{F_{\max}}{F_{\text{avg}}}=\frac{\max_i(F_i)}{\left(\sum_i F_i\right)/n}=\frac{\max_i(F_i)}{F_H/n}$$

is the ratio of the maximum fund held by any single honest node to the average fund across all honest nodes in the network.

Proof. Let P be the random path selected by the random-path selection algorithm. The OM sent on P fails if and only if at least one of the following events occurs: 1) event $\mathcal{E}_{drp}$: the path P contains an adversarial node, in which case the adversarial node will drop the message; 2) event $\mathcal{E}_{sat}$: the path P contains an honest link saturated by the adversary[1], in which case the OM is dropped by the receiver of the link.

In the following, we find an upper bound on the probability for each of the above two events. We then simply add these probabilities to get an upper bound on the probability of OM failure.

To select the random path P, the random-path selection algorithm chooses $(l-1)$ nodes at random in $(l-1)$ rounds, where in each round, the probability that a given node is selected is proportional to its fund. Therefore, $\Pr(\mathcal{E}_{drp})$, the probability that the random-path selection algorithm selects a node that belongs to the adversary is simply equal to

$$\Pr(\mathcal{E}_{drp})=1-(1-\gamma)^{l-1}.$$

Next, we establish an upper bound on $\Pr(\mathcal{E}_{\text{sat}})$.

The adversary can establish up to n connections to all the n honest nodes, where each connection can push a rate of at most $\alpha\cdot F_A$. Moreover, each connection can engage at most $(L-1)$ honest links (the first link is not an honest link as it is incident to an adversarial node, hence $(L-1)$ instead of L). Therefore, the sum of all the rates the adversary can impose on honest links in the network denoted $\mathfrak{b}$, is at most

$$\mathfrak{b}\leq\alpha\cdot F_A\cdot(L-1)\cdot n \tag{1}$$

The first and last links on path P do not drop the OM, even when saturated. This is because the source node prioritizes forwarding its own OMs, and the destination node always accepts OMs destined for itself. Therefore, in analyzing $\Pr(\mathcal{E}_{\text{sat}})$, we only consider the other $l-2$ links on the path P.

[1]A link other than the first and the last links as those links belong to, respectively, the source and destination and do not drop the OM message.

The probability that the kth link, $1<k<l$, on the path P is a saturated honest link is

$$\sum_{i,j} f_i\cdot f_j\cdot I_{i,j}, \tag{2}$$

where $f_i=\frac{F_i}{F_A+F_H}$ is the probability that node U_i is selected at any position by the random-path selection algorithm. The term $f_i\cdot f_j$ represents the probability that the kth link is (U_i,U_j), and $I_{i,j}$ is an indicator variable equal to 1 if the adversary has saturated the link from U_i to U_j, and 0 otherwise. To set $I_{i,j}$ to one, the adversary needs to push OMs at a rate of at least αF_i from U_i to U_j. Since the adversary's maximum budget is $\mathfrak{b}$, we must then have

$$\sum_{i,j}(\alpha\cdot F_i)\cdot I_{i,j}\leq\mathfrak{b} \tag{3}$$

Thus, we get

$$\begin{aligned}
\Pr(\mathcal{E}_{\text{sat}}) &\overset{by(2)\text{ and union bound}}{\leq} (l-2)\cdot\left(\sum_{i,j} f_i\cdot f_j\cdot I_{i,j}\right)\\
&\overset{f_{max}=\max_i(f_i)}{\leq} (l-2)\cdot\left(\sum_{i,j} f_i\cdot f_{max}\cdot I_{i,j}\right)\\
&= (l-2)\cdot f_{max}\cdot\left(\sum_{i,j} f_i\cdot I_{i,j}\right)\\
&= (l-2)\cdot f_{max}\cdot\left(\sum_{i,j}\frac{F_i}{F_A+F_H}\cdot I_{i,j}\right)\\
&= (l-2)\cdot\frac{f_{max}}{F_A+F_H}\cdot\left(\sum_{i,j} F_i\cdot I_{i,j}\right)\\
&= (l-2)\cdot\frac{f_{max}}{F_A+F_H}\cdot\frac{1}{\alpha}\cdot\left(\sum_{i,j}(\alpha\cdot F_i)\cdot I_{i,j}\right)\\
&\overset{(3)}{\leq} (l-2)\cdot\frac{f_{max}}{F_A+F_H}\cdot\frac{1}{\alpha}\cdot\mathfrak{b}
\end{aligned} \tag{4}$$

Therefore, substituting the expression for $\mathfrak{b}$ from (1) yields

$$\begin{aligned}
\Pr(\mathcal{E}_{\text{sat}}) &\overset{(1)}{\leq} (l-2)\cdot\frac{f_{max}}{F_A+F_H}\cdot\frac{1}{\alpha}\cdot(\alpha\cdot F_A\cdot(L-1)\cdot n)\\
&= (l-2)\cdot\frac{f_{max}}{F_A+F_H}\cdot(F_A\cdot(L-1)\cdot n)\\
&= (L-1)\cdot(l-2)\cdot\frac{F_A}{F_A+F_H}\cdot(f_{max}\cdot n)\\
&= (L-1)\cdot(l-2)\cdot\gamma\cdot(f_{max}\cdot n)\\
&= (L-1)\cdot(l-2)\cdot\gamma\cdot\left(\frac{F_{max}}{F_A+F_H}\cdot n\right)\\
&= (L-1)\cdot(l-2)\cdot\gamma\cdot\left(\frac{F_H}{F_H}\cdot\frac{F_{max}}{F_A+F_H}\cdot n\right)\\
&= (L-1)\cdot(l-2)\cdot\gamma\cdot\left(\frac{F_H}{F_A+F_H}\cdot\frac{F_{max}}{F_H/n}\right)\\
&= (L-1)\cdot(l-2)\cdot\gamma\cdot(1-\gamma)\cdot\delta
\end{aligned} \tag{5}$$

Finally, by a union bound, we get that the probability that the OM fails (over the choice of random path selection by the random-path selection algorithm) is at most equal to

$$\begin{aligned}&\Pr(\mathcal{E}_{drp}) + \Pr(\mathcal{E}_{\text{sat}})\\ &\leq \left[1-(1-\gamma)^{l-1}\right] + \left[(L-1)\cdot(l-2)\cdot\gamma\cdot(1-\gamma)\cdot\delta\right]\end{aligned} \tag{6}$$

thus

$$\begin{aligned}&\Pr(\mathcal{E}_{drp}) + \Pr(\mathcal{E}_{\text{sat}})\\ &\leq \left[1-(1-\gamma)^{l-1}\right] + \left[(L-1)\cdot(l-2)\cdot\gamma\cdot(1-\gamma)\cdot\delta\right]\\ &\leq \left[(l-1)\cdot\gamma\right] + \left[(L-1)\cdot(l-2)\cdot\gamma\cdot(1-\gamma)\cdot\delta\right]\\ &\leq \left[(L-1)\cdot\gamma\right] + \left[(L-1)\cdot(l-2)\cdot\gamma\cdot(1-\gamma)\cdot\delta\right]\\ &\leq \left[(L-1)\cdot\gamma\right] + \left[(L-1)\cdot(l-2)\cdot\gamma\cdot\delta\right]\\ &= (L-1)\cdot\gamma\cdot\left[1+(l-2)\cdot\delta\right]\\ &= (L-1)\cdot\gamma\cdot\left[1-\delta+(l-1)\cdot\delta\right]\\ &\leq (L-1)\cdot(l-1)\cdot\gamma\cdot\delta\end{aligned}$$

where the last inequality is because $\delta \geq 1$ as the maximum value of a set of numbers is not less than their average, hence $\delta = \frac{\max_{i=1}^{n}(F_i)}{\sum_{i=1}^{n} F_i/n} \geq 1$.

□

Sybil resistance. By Theorem 1, the power of the adversary (the probability of OM failure) is at most $(L-1)(l-1)\,\gamma\,\delta$, where $\gamma = F_A/(F_A+F_H)$ is the ratio between the adversary's total funds F_A and the total funds in the network, and $\delta = F_{\max}/F_{\text{avg}}$ captures the skew in the *honest* fund distribution (with $F_{\max}$ and F_{avg} taken over honest nodes only). Thus the adversary's effect is governed by its aggregate share of network capital, not by how many Sybil nodes it creates: splitting the same F_A over more nodes leaves γ unchanged. The ratio of the adversary's funds to any single honest node's balance (including the largest one) does not appear as an independent parameter; its influence is already reflected indirectly through γ and δ.

Impact of Fund Distribution. Consider two networks with the same total honest funds F_H and the same total adversarial funds F_A, but with different distributions of honest funds. Assume that both networks use the same random-path selection algorithm, with identical parameters l and L. A natural question is which network provides greater resilience against OM failures.

Theorem 1 indicates that the distribution of funds can significantly influence the upper bound on OM failure probability. Given that both networks have the same total honest and adversarial funds, they share the same value for γ as well as identical values for l and L.

The critical difference between the two networks in the context of Theorem 1 is the parameter $\delta = \frac{F_{\max}}{F_{\text{avg}}}$, which captures the disparity in fund distribution. In a network with a uniform fund distribution, this ratio is minimized at $\delta = 1$, reflecting the fact that each honest node holds an equal share of the total funds. In contrast, for non-uniform distributions, $\delta > 1$ since the largest honest node holds more than the average fund amount.

The following corollary states that the bound established in Theorem 1 increases linearly with δ, highlighting the importance of fund distribution in network resilience.

Corollary 1. *Let the total honest funds F_H, the total adversarial funds F_A, and the random-path selection algorithm parameters l and L be fixed. Then, the upper bound on the OM failure probability given in Theorem 1 is a linear function of δ and achieves its minimum at $\delta = 1$.*

Proof. The bound in Theorem 1 can be expressed as

$$B(\delta) = C_1 + C_2 \cdot \delta,$$

where C_1 and C_2 are constants dependent only on the fixed parameters F_A, F_H, l, and L. Thus, $B(\delta)$ increases linearly with δ. Since, by definition, $\delta \geq 1$, the minimum occurs at $\delta = 1$. □

To further illustrate the impact of fund distribution on network resilience, the following example compares the adversary's influence in two scenarios that differ only in how funds are distributed among honest nodes. In the first scenario, the total funds are uniformly allocated among all honest nodes; in the second scenario, the same total funds are allocated non-uniformly.

Example 1. *Consider a network with n honest nodes and a fixed total fund F_H. Assume that the adversary's total fund equals the average fund of honest nodes, that is, $F_A = F_{\text{avg}} = F_H/n$. Let $L = l = 3$. In the first scenario, the total fund of F_H is uniformly distributed among the honest nodes, such that each node receives $F_H/n = F_{\text{avg}}$. For this scenario, we have $\gamma_1 = \frac{F_A}{F_A+F_H} = \frac{F_H/n}{F_H/n+F_H} = \frac{1}{n+1}$ and $\delta_1 = \frac{F_{\max}}{F_{\text{avg}}} = 1$. Thus, by Theorem 1, the probability of failure is given by:*

$$\begin{aligned}\Pr(\textit{fail}_1) &= \mathcal{O}(L\cdot l\cdot\gamma_1\cdot\delta_1)\\ &= \mathcal{O}\left(\frac{1}{n}\right).\end{aligned} \tag{7}$$

In the second scenario, suppose the total fund of F_H is distributed non-uniformly among honest nodes. Specifically, the first $n-1$ honest nodes receive identical funds of f, but the n'th node receives a fund of $m\cdot f$, where $1 < m < n$ is a number.[2] *For this scenario, we get*

$$\gamma_2 = \frac{F_A}{F_A+F_H} = \frac{F_H/n}{F_H/n+F_H} = \frac{1}{n+1} = \gamma_1,$$

and

$$\delta_2 = \frac{F_{\max}}{F_{\text{avg}}} = \frac{m\cdot f}{\frac{m\cdot f+(n-1)f}{n}} = \frac{m\cdot n}{m+n-1}$$

Therefore, by Theorem 1, we have:

$$\begin{aligned}\Pr(\textit{fail}_2) &= \mathcal{O}(L\cdot l\cdot\gamma_2\cdot\delta_2)\\ &= \mathcal{O}(\gamma_2\cdot\delta_2)\\ &= \mathcal{O}\left(\frac{1}{n+1}\cdot\frac{m\cdot n}{m+n-1}\right)\\ &= \mathcal{O}\left(\frac{m}{n}\right).\end{aligned} \tag{8}$$

[2]The value of f is set to $\frac{F_H}{n+m-1}$, ensuring that the total funds of all honest nodes sum to F_H.

Note that $\gamma_2 = \gamma_1$, but $\delta_2 = \Theta(m) \cdot \delta_1$. Thus, according to Theorem 1, the probability of failure in the second scenario is expected to be a factor of $\Theta(m)$ higher than in the first. The adversary can exploit this heightened risk to significantly increase the probability of failure. To demonstrate this, consider the adversary sends OMs at a rate of $\alpha \cdot F_A = \alpha \cdot F_{\text{avg}}$ to each honest node U_i, where $1 \le i < n$, and directs them to forward these OMs to node U_n, effectively saturating all links to U_n.

Assume two honest nodes, U_s and U_d, with U_s using the random-path selection algorithm to select a random path $P = (U_s, U_i, U_j, U_d)$ of length $l = 3$ to send an OM to U_d through random nodes U_i and U_j. The link (U_i, U_j) is saturated iff $j = n$ and $1 \le i \le n-1$. Therefore, the probability that the link (U_i, U_j) is saturated equals to

$$\frac{(n-1)F_{\text{avg}}}{n \cdot F_{\text{avg}}} \cdot \frac{F_{\text{avg}}}{n \cdot F_{\text{avg}}} = \frac{n-1}{n^2} = \Theta\left(\frac{1}{n}\right),$$

in the first scenario, while the equivalent probability in the second scenario is

$$\frac{(n-1)\cdot f}{(n-1)\cdot f + m \cdot f} \cdot \frac{m \cdot f}{(n-1)\cdot f + m \cdot f} = \frac{(n-1)\cdot m}{(n+m-1)^2} = \Theta\left(\frac{m}{n}\right).$$

These probabilities are asymptotically optimal as they match the upper bound established by Theorem 1, with the latter being a factor of $\Theta(m)$ larger than the former, illustrating the significant impact of a non-uniform fund distribution.

A tighter bound. We can further tighten the bound established in Theorem 1. The idea to do this is by making the second inequality in (4) tighter. By doing so, Theorem 2 replaces $\delta = \frac{F_{\max}}{F_{\text{avg}}}$ in Example 1 with $\check{\delta} = \frac{F^{(k)}_{\max}}{F_{\text{avg}}}$, where $F^{(k)}_{\max}$ is the average funds of the top k honest nodes, for some k defined in Theorem 2. Since clearly $F^{(k)}_{\max} \le F_{\max}$, we get that $\check{\delta} \le \delta$, hence we get a tighter upper bound in Theorem 2.

Theorem 2. *Let L denote the maximum number of hops an OM can travel, and $l \le L$ denote the OM path-length parameter used in the random-path selection algorithm. Let F_A be the total funds of the adversary, and $F_H = \sum_{i=1}^{n} F_i$ be the total fund of honest nodes, where F_i denotes the fund of honest node U_i, and n is the total number of honest nodes in the network.*
Then, for every pair of distinct honest nodes U_s and U_d, the probability that an OM message, sent on a random path chosen by the random-path selection algorithm, from U_s to U_d fails is at most

$$1 - (1-\gamma)^{l-1} + (L-1)\cdot(l-2)\cdot\gamma\cdot\check{\delta} \le (L-1)\cdot(l-1)\cdot\check{\delta}\cdot\gamma$$

where

$$\gamma = \frac{F_A}{F_A + F_H}$$

is the fraction of funds in the entire network that belongs to the adversary, and

$$\check{\delta} = \frac{F^{(k)}_{\max}}{F_{\text{avg}}}$$

is the ratio of the average fund held by the top k honest nodes to the average fund across all honest nodes in the network, and

$$k = \min\left\{n,\ \max\left\{1, \lfloor \frac{(L-1)F_A}{F_{\text{avg}}} \rfloor\right\}\right\}.$$

Proof. Let $T = F_A + F_H$, and assume without loss of generality that

$$F_1 \ge F_2 \ge \cdots \ge F_n.$$

Let $\overline{\mathfrak{b}} = \alpha F_A (L-1) n$ be the upper bound on the adversary's aggregate saturation budget. For each potential target U_j, define $c_j = \sum_i \alpha F_i I_{i,j}$. Thus, c_j is the portion of the saturation budget allocated to incoming links of U_j. We have

$$0 \le c_j \le \alpha F_H \quad \text{and} \quad \sum_j c_j \le \mathfrak{b} \le \overline{\mathfrak{b}}.$$

For one interior position on the path,

$$\sum_{i,j} f_i f_j I_{i,j} = \frac{1}{T^2}\sum_{i,j} F_i F_j I_{i,j} = \frac{1}{\alpha T^2}\sum_j F_j c_j.$$

Because $F_1 \ge F_2 \ge \cdots \ge F_n$, this expression is maximized by allocating the budget to targets in decreasing order of their weights. Let

$$x = \frac{\overline{\mathfrak{b}}}{\alpha F_H} = \frac{(L-1)F_A}{F_{\text{avg}}}.$$

If $1 \le x < n$, write

$$x = k + r, \qquad k = \lfloor x \rfloor, \qquad 0 \le r < 1.$$

The first k targets can receive a full allocation of αF_H, and the next target can receive at most a fraction r of that amount. Therefore,

$$\begin{aligned}\sum_j F_j c_j &\le \alpha F_H \left(\sum_{j=1}^{k} F_j + r F_{k+1}\right) \\ &\le \alpha F_H (k+r) F^{(k)}_{\max} \\ &= \overline{\mathfrak{b}} F^{(k)}_{\max}.\end{aligned}$$

The same inequality is immediate when $x < 1$ by taking $k = 1$, and when $x \ge n$ by taking $k = n$.

It follows that

$$\begin{aligned}\Pr(\mathcal{E}_{\text{sat}}) &\le (l-2)\frac{\overline{\mathfrak{b}} F^{(k)}_{\max}}{\alpha (F_A + F_H)^2} \\ &= (l-2)\frac{\alpha F_A (L-1) n F^{(k)}_{\max}}{\alpha (F_A + F_H)^2} \\ &= (L-1)(l-2)\frac{F_A}{F_A+F_H}\frac{F_H}{F_A+F_H}\frac{F^{(k)}_{\max}}{F_H/n} \\ &= (L-1)(l-2)\gamma(1-\gamma)\check{\delta}.\end{aligned}$$

Combining this result with

$$\Pr(\mathcal{E}_{\text{drp}}) = 1 - (1-\gamma)^{l-1}$$

gives the first bound. The second follows from

$$1 - (1-\gamma)^{l-1} \le (l-1)\gamma,$$

$l \le L$, $1 - \gamma \le 1$, and $\check{\delta} \ge 1$. □

Adversarial strategy. Theorems 1 and 2 bound the damage that any adversary with a fixed total fund F_A can inflict, yet some strategies can drive the failure probability much closer to these bounds than others. We refer to the following as the *optimal strategy*, because it maximizes—per unit of fund—the probability that a specific directed communication link on the path is saturated.

Assume the adversary seeks to maximize the likelihood that the h-th hop of an OM path, where $1 < h < l$, becomes saturated. Since the random-path selection algorithm chooses U_i and U_j independently—selecting U_i with probability f_i and U_j with probability f_j—the probability that the h-th hop connects these two nodes is $f_i f_j$. Saturating that link costs $\alpha F_i = cf_i$, where $c = \alpha\,(F_A + F_H)$ is a constant. Hence, a budget outlay of cf_i buys a success probability of $f_i f_j$. To maximize the gain–cost ratio, the adversary then must choose U_j with the largest f_j—that is, the honest node holding the greatest funds. Once all links incident to this richest node are saturated, any remaining budget is directed to links incident to the next-richest node, and so on. Thus, to maximize the OM failure probability, the adversary should systematically saturate directed communication links into the highest-weight honest nodes.

In Section VI, we evaluate three adversarial strategies. The *random-target adversary* distributes the attack traffic available through each direct connection among randomly chosen targets, whereas the *top-target adversary* directs this traffic toward the highest-weight targets first. Both are per-connection stress tests that respect the forwarding-rate limit on each direct adversarial connection but are not constrained by the payment-channel topology. The *strong optimal adversary* may allocate its aggregate budget arbitrarily and serves as a conservative benchmark. For the evaluated setting $l = L = 3$, the top-target and strong optimal adversaries produce essentially identical failure probabilities. Over the full network, their results also closely match the upper bound in Theorem 2; over the selected subsets, they remain below the bound in Theorem 3. The random-target adversary is substantially less effective.

B. Attacking a Victim Node

In this attack scenario, the adversary aims to prevent nodes from reaching a victim node W. We demonstrate that the probability of any honest node U's message to W being dropped before reaching W is bounded by the limits specified in Theorem 1 and Theorem 2.

Let P denote the random path selected by the random-path selection algorithm. As before, the OM sent on P fails if at least one of the following events occurs: 1) Event $\mathcal{E}_{drp}$: The path P contains an adversarial node, which will drop the message; 2) Event $\mathcal{E}_{sat}$: The path P contains an honest link saturated by the adversary, leading to the OM being dropped by the link's receiver.

The probability that the random-path selection algorithm selects a path containing an adversarial node is given by

$$\Pr(\mathcal{E}_{drp}) = 1 - (1-\gamma)^{l-1},$$

consistent with previous results.

To analyze $\Pr(\mathcal{E}_{\text{sat}})$, note that according to the forwarding rules, if the last node on P is honest, then the link from this node to W will never drop a message destined for W, even if the link is saturated. Moreover, all nodes on path P are selected by the random-path selection algorithm independently of the destination node W. Therefore, $\Pr(\mathcal{E}_{\text{sat}})$ is independent of the destination node choice. This implies that in analyzing $\Pr(\mathcal{E}_{\text{sat}})$, we can consider W to be any node in the network, and hence, the analysis of $\Pr(\mathcal{E}_{\text{sat}})$ mirrors those presented in Theorem 1 and Theorem 2.

IV. Our Proposed Method

In this section, we propose restricting forwarding-node selection to a carefully selected subset of nodes to improve OM resilience in networks where funds are not uniformly distributed. We then demonstrate that this method can significantly tighten the upper bound established in Theorem 1—specifically in networks with $\delta > 1$. This consideration is particularly relevant in practice, as δ ranges from 442.500 to 740.392 across the three LN snapshots used in our evaluation.

In a nutshell, our proposed method restricts the random-path selection algorithm to a subset S of nodes when selecting forwarding nodes. Importantly, the funds of each node in this subset remain defined by their total funds in the entire network—not just their share within S.

The following theorem establishes an upper bound on the probability of OM failure under this method, expressed in terms of the total and maximum funds within S. A key feature of this bound is that it does not depend on unknown parameters such as the total fund of honest nodes within the subset (since honest nodes within the set S are unknown). Leveraging this feature, we present in the remainder of this section a polynomial-time algorithm (Algorithm 1) that identifies an optimal subset S—that is, one that minimizes the derived upper bound on OM failure probability.

Remark 1. *In establishing the upper bound, we assume that the adversary can adopt any arbitrary strategy, constrained only by a fixed total fund budget of F_A. Specifically, the adversary may allocate these funds to saturate any set of links in the network, provided the total allocation does not exceed this budget. Additionally, we adopt an* adaptive-adversary *model, where the adversary selects its corrupted nodes after the forwarding subset S has been fixed, with full knowledge of both S and the modified random-path selection algorithm, which restricts path selection to nodes within the set S.*

We note that this modification does indeed affect the adversary's optimal strategy, as the new mechanism restricts message forwarding to a subset S rather than the entire network. While this may increase the adversary's relative presence within S—and thus make it appear more powerful—our analysis directly incorporates this worst-case condition. Despite this, the resilience of OM forwarding improves significantly. Conceptually, this is because the subset S can be chosen to yield a substantially smaller δ value, which plays a critical role in the failure probability bound.

Theorem 3. *Suppose we restrict the selection of forwarding nodes in the random-path selection algorithm to a subset S*

of nodes within the network. Then, for every pair of honest nodes U_s and U_d, the probability that an OM message, sent along a randomly chosen path by the random-path selection algorithm from U_s to U_d, fails is at most

$$(l-1)\cdot F_A\cdot\left(\frac{1}{F_{\text{tot}}^{(s)}}+(L-1)\cdot\frac{|S|\cdot F_{\max}^{(s)}}{\left(F_{\text{tot}}^{(s)}\right)^2}\right) \tag{9}$$

where F_A represents the total funds of the adversary in the entire network, $F_{\max}^{(s)}$ is the maximum fund held by any single node in the subset S, and $F_{\text{tot}}^{(s)}$ is the total fund of all nodes within S.

Proof. Let P be the random path selected by the random-path selection algorithm when forwarding nodes are restricted to the subset S. The OM message sent on P fails if and only if at least one of the following events occurs: 1) Event $\mathcal{E}_{drp}$: the path P includes an adversarial node, which consequently drops the message. 2) Event $\mathcal{E}_{sat}$: the path P features an honest link that is saturated by the adversary, leading to the OM message being dropped by the receiver of that link. This is the same failure condition as before, with the only difference being that the forwarding nodes on P are now restricted to the set S.

To construct the random path P, the random-path selection algorithm selects $(l-1)$ intermediate nodes from the subset S, choosing one node in each of the $(l-1)$ rounds. In each round, the probability of selecting a specific node in S is proportional to its fund in the entire network. Consequently, the probability $\Pr(\mathcal{E}_{drp})$, that the random-path selection algorithm selects an adversarial node, is given by

$$\begin{aligned}\Pr(\mathcal{E}_{drp}) &= 1-\left(1-\frac{F'_A}{F_{\text{tot}}^{(s)}}\right)^{l-1}\\ &\le (l-1)\cdot\frac{F_A}{F_{\text{tot}}^{(s)}},\end{aligned} \tag{10}$$

where $F'_A \le F_A$ represents the total fund of the adversary's nodes within the subset S. Since our model assumes a strong *adaptive-adversary* setting—where the adversary selects its corrupted nodes *after* the forwarding subset S has been fixed—we take $F'_A = F_A$.

Let $F_1, F_2, \ldots, F_h$ denote the funds of honest nodes $U_1, U_2, \ldots, U_h$ in S, where h denotes the number of honest nodes in S. Note that only the nodes in S forward OMs. The adversary is capable of establishing up to h connections with these honest nodes, each connection pushing a rate of at most $\alpha\cdot F_A$. Furthermore, each connection can engage up to $(L-1)$ honest links. Consequently, the total rate that the adversary can impose on links between honest nodes in S, denoted by $\mathfrak{b}^{(s)}$, is bounded by

$$\begin{aligned}\mathfrak{b}^{(s)} &\le \alpha\cdot F_A\cdot(L-1)\cdot h\\ &\le \alpha\cdot F_A\cdot(L-1)\cdot|S|\end{aligned} \tag{11}$$

The probability that the kth link, $1<k<l$, on the path P is a saturated honest link is

$$\sum_{1\le i,j\le h} f_i\cdot f_j\cdot I_{i,j}, \tag{12}$$

where $f_i=\frac{F_i}{F_{\text{tot}}^{(s)}}$, $f_j=\frac{F_j}{F_{\text{tot}}^{(s)}}$, $f_i\cdot f_j$ is the probability that the kth link, $1<k<l$, is (U_i, U_j), and $I_{i,j}$ is a binary variable equal to one if the adversary has saturated (U_i, U_j), the link from U_i to U_j; and equal to zero otherwise. To set $I_{i,j}$ to one, the adversary needs to push OMs at a rate of at least αF_i from U_i to U_j. Given the adversary's maximum budget is $\mathfrak{b}^{(s)}$, we must then have

$$\sum_{1\le i,j\le h}(\alpha\cdot F_i)\cdot I_{i,j}\le\mathfrak{b}^{(s)}. \tag{13}$$

Thus, we get

$$\begin{aligned}\Pr(\mathcal{E}_{\text{sat}}) &\overset{by(12)\text{ and union bound}}{\le} (l-2)\cdot\left(\sum_{1\le i,j\le h} f_i\cdot f_j\cdot I_{i,j}\right)\\ &\overset{f_{max}^{(s)}=\max_{U_i\in S}(f_i)}{\le} (l-2)\cdot\left(\sum_{1\le i,j\le h} f_i\cdot f_{max}^{(s)}\cdot I_{i,j}\right)\\ &= (l-2)\cdot f_{max}^{(s)}\cdot\left(\sum_{1\le i,j\le h} f_i\cdot I_{i,j}\right)\\ &= (l-2)\cdot f_{max}^{(s)}\cdot\left(\sum_{1\le i,j\le h}\frac{F_i}{F_{tot}^{(s)}}\cdot I_{i,j}\right)\\ &= (l-2)\cdot\frac{f_{max}^{(s)}}{F_{tot}^{(s)}}\cdot\left(\sum_{1\le i,j\le h} F_i\cdot I_{i,j}\right)\\ &= (l-2)\cdot\frac{f_{max}^{(s)}}{F_{tot}^{(s)}}\cdot\frac{1}{\alpha}\cdot\left(\sum_{1\le i,j\le h}(\alpha\cdot F_i)\cdot I_{i,j}\right)\\ &\overset{(13)}{\le} (l-2)\cdot\frac{f_{max}^{(s)}}{F_{tot}^{(s)}}\cdot\frac{1}{\alpha}\cdot\mathfrak{b}^{(s)}\\ &\overset{(11)}{\le} (l-2)\cdot\frac{f_{max}^{(s)}}{F_{tot}^{(s)}}\cdot\frac{1}{\alpha}\cdot(\alpha\cdot F_A\cdot(L-1)\cdot|S|)\\ &= (l-2)\cdot\frac{f_{max}^{(s)}}{F_{tot}^{(s)}}\cdot(F_A\cdot(L-1)\cdot|S|)\\ &= (L-1)\cdot(l-2)\cdot\frac{F_A}{F_{tot}^{(s)}}\cdot\left(\frac{F_{max}^{(s)}\cdot|S|}{F_{tot}^{(s)}}\right)\\ &= (L-1)\cdot(l-2)\cdot\frac{|S|\cdot F_A\cdot F_{max}^{(s)}}{\left(F_{tot}^{(s)}\right)^2}.\end{aligned} \tag{14}$$

Note that the above bound is particularly strong, as the inequality (13) used in its derivation holds regardless of the adversary's strategy, and remains valid even if the adversary is permitted to saturate any set of links, subject only to the constraint imposed by its total fund F_A.

Finally, by (14), (10) and a union bound, we get that the probability that the OM fails (over the choice of random path selection by the random-path selection algorithm) is at most equal to

$$
\begin{aligned}
&\Pr(\mathcal{E}_{drp}) + \Pr(\mathcal{E}_{\text{sat}}) \\
&\leq \left((l-1)\cdot\frac{F_A}{F_{\text{tot}}^{(s)}}\right) + \left((L-1)\cdot(l-2)\cdot\frac{|S|\cdot F_A\cdot F_{\max}^{(s)}}{\left(F_{\text{tot}}^{(s)}\right)^2}\right) \\
&\leq \left((l-1)\cdot\frac{F_A}{F_{\text{tot}}^{(s)}}\right) + \left((L-1)\cdot(l-1)\cdot\frac{|S|\cdot F_A\cdot F_{\max}^{(s)}}{\left(F_{\text{tot}}^{(s)}\right)^2}\right) \\
&\leq (l-1)\cdot F_A\cdot\left(\frac{1}{F_{\text{tot}}^{(s)}} + (L-1)\cdot\frac{|S|\cdot F_{\max}^{(s)}}{\left(F_{\text{tot}}^{(s)}\right)^2}\right)
\end{aligned}
\tag{15}
$$

□

Optimum set. We define a subset $S \subseteq V$ as optimum if it minimizes the upper bound (9) of Theorem 3. Note that the leading coefficients $(l-1)$ and F_A in (9) are independent of the choice of the set S. Thus, a set S is optimum if it is a member of

$$
\mathcal{S} = \arg\min_{S\subseteq V}\left(\frac{1}{F_{\text{tot}}^{(s)}} + (L-1)\cdot\frac{|S|\cdot F_{\max}^{(s)}}{\left(F_{\text{tot}}^{(s)}\right)^2}\right) \tag{16}
$$

Although the set V has an exponential number of subsets S, Theorem 4 demonstrates that a set in $\mathcal{S}$ can be found in quadratic time. We use this set in the simulations to show that LN can improve its resilience against the adversary by restricting forwarding-node selection to a carefully selected subset.

Lemma 1. *Consider a network with m users $U_1, U_2, \ldots, U_m$. Let F_i denote the funds of node U_i in the network. Suppose $F_1 \leq F_2 \leq \ldots \leq F_m$.*
Then there exists a set $S \in \mathcal{S}$ such that $S = \{U_k | k \in [i,j]\}$ for some integers $1 \leq i \leq j \leq m$.

Proof. Let us define the *spread* of a set as the greatest difference between the subscript indices of any two nodes in the set. For instance, the spread of the set $\{U_2, U_5, U_7, U_8\}$ is $8-2=6$. Let S be a set in $\mathcal{S}$ with the minimum spread. We show that $S = \{U_k | k \in [i,j]\}$ for some integers $1 \leq i \leq j \leq m$. Let U_i and U_j be the nodes in S with the maximum difference between their indices. Towards a contradiction, suppose that there is an integer $i < k < j$ such that $U_k \notin S$. Let $S^\dagger = (S \cup \{U_k\}) \setminus \{U_i\}$, that is $S^\dagger$ is formed from S by adding U_k to and removing U_i from S. We have $|S^\dagger| = |S|$, $F_{\max}^{(s^\dagger)} = F_{\max}^{(s)}$ and $F_{\text{tot}}^{(s)} \leq F_{\text{tot}}^{(s^\dagger)}$ as $F_i \leq F_k$. Therefore, by (16), we get

$$
S \in \mathcal{S} \Longrightarrow S^\dagger \in \mathcal{S}.
$$

This is a contradiction since $S^\dagger$ has a strictly smaller spread than S, but S was assumed to have the smallest spread among the sets in $\mathcal{S}$.

□

Theorem 4. *There exists a quadratic-time algorithm (Algorithm 1) for finding a set in $\mathcal{S}$.*

Proof. By Lemma 1, there is a set $S \in \mathcal{S}$ such that $S = \{U_k | k \in [i,j]\}$ for some integers $1 \leq i \leq j \leq n$. There are quadratically many sets of this form. Therefore, as described in Algorithm 1, one that belongs to $\mathcal{S}$ can be identified in quadratic time. □

Algorithm 1: Finds an element S of the set $\mathcal{S}$ defined in (16).

Input: Nodes $U_1, \ldots, U_n$, their public-capacity weights $F_1, \ldots, F_n$, and the maximum OM travel distance L
Output: A set $S \in \mathcal{S}$

Sort and relabel the nodes such that $F_1 \leq F_2 \leq \cdots \leq F_n$, breaking ties by node public key;
$minValue \leftarrow \infty$;
$bestStart \leftarrow 1$;
$bestEnd \leftarrow 1$;
for $start \leftarrow 1$ **to** n **do**
  $sSum \leftarrow 0$;
  for $end \leftarrow start$ **to** n **do**
    $sSum \leftarrow sSum + F_{end}$;
    $sLen \leftarrow end - start + 1$;
    $sMax \leftarrow F_{end}$ `// The weights are sorted`
    $currentValue \leftarrow \frac{1}{sSum} + (L-1)\frac{sLen \cdot sMax}{(sSum)^2}$;
    if $currentValue < minValue$ **then**
      $minValue \leftarrow currentValue$;
      $bestStart \leftarrow start$;
      $bestEnd \leftarrow end$;
    end if
  end for
end for
$S \leftarrow \{U_k \mid bestStart \leq k \leq bestEnd\}$;
return S;

Deployment considerations. The forwarding subset is a sender-side path-selection policy and requires no new consensus: relays need not know or agree on the sender's subset. Each sender computes it locally from its validated gossip view. With the same view and tie-breaking rule, the deterministic algorithm returns the same subset. It may be recomputed periodically or after a material change in the public channel graph.

Some forwarding concentration is inherent to the public-capacity-weighted baseline, since relay-selection probabilities are already non-uniform when the full network is used. Restricting the forwarding pool to S introduces additional concentration by renormalizing these probabilities; Section VI quantifies this incremental traffic and availability burden. The subset returned by Algorithm 1 optimizes the resilience objec-

tive and is not a mandatory deployment size; an implementation may instead impose a minimum cardinality.

Under churn or inconsistent gossip views, senders may obtain different subsets. They can omit unavailable relays, resample, recompute later, or temporarily use the full forwarding pool. This does not affect OM validity, although the common-subset bound may not apply until views converge; our numerical results assume a common snapshot. The subset is also publicly computable by the adversary, which is already accounted for in Theorem 3.

V. Privacy Analysis

Let $\mathcal{A}$ be any probabilistic polynomial-time (PPT) adversary that corrupts a node set $\mathcal{C} \subseteq V$. A corrupted node can see the full contents of any OM it processes (ciphertexts remain unintelligible, but headers and path information are visible). Global traffic analysis, timing attacks, and breaks of end-to-end encryption are out of scope.

Let $P = (U_0, \ldots, U_\ell)$ denote the OM path chosen by the random-path selection algorithm, where U_0 is the first hop after the sender and U_ℓ is the hop that forwards to the receiver. Define the event

$$\mathcal{H} \;:=\; \big(U_0 \notin \mathcal{C} \;\wedge\; U_\ell \notin \mathcal{C}\big),$$

i.e. $\mathcal{H}$ occurs iff *both* endpoints of P are honest. We first show that perfect endpoint privacy is guaranteed when $\mathcal{H}$ holds. Then, we calculate the probability that $\mathcal{H}$ does not hold.

Definition 2 (Conditional Endpoint Anonymity)**.** *Let S (resp. R) be the random variable denoting the sender (resp. receiver) of an OM, and let $\mathsf{View}_{\mathcal{A}}(s, r)$ be the adversary's view when sender s transmits an OM to receiver r. We say π_{OM} achieves* conditional sender (resp. receiver) anonymity *if, for any PPT adversary $\mathcal{A}$, any two honest senders $s_0, s_1 \notin \mathcal{C}$ (resp. receivers $r_0, r_1 \notin \mathcal{C}$), and any honest receiver r (resp. sender s), the distributions*

$$\mathsf{View}_{\mathcal{A}}(s_0, r) \;\big|\; \mathcal{H} \quad \textit{and} \quad \mathsf{View}_{\mathcal{A}}(s_1, r) \;\big|\; \mathcal{H}$$

(resp. the analogous pair for receivers) are identical. Equivalently,

$$I(S; P \mid \mathcal{H}) = 0 \quad \textit{and} \quad I(R; P \mid \mathcal{H}) = 0.$$

Theorem 5 (Conditional Endpoint Anonymity)**.** *For any PPT adversary $\mathcal{A}$ that may corrupt an arbitrary set of* interior *nodes, protocol π_{OM} satisfies Definition 2.*

Proof. We prove sender anonymity; the receiver case is symmetric. The random-path selection algorithm samples P *without consulting* either endpoint. Hence the distribution of P is independent of the sender: $\Pr[P \mid S = s_i, \mathcal{H}] = \Pr[P \mid \mathcal{H}]$ for any honest senders s_i. Therefore $H(P \mid S, \mathcal{H}) = H(P \mid \mathcal{H})$, yielding $I(S; P \mid \mathcal{H}) = 0$. The adversary's view is a function of P and local randomness only, so the two views in Definition 2 are identically distributed. □

Corollary 2 (Perfect Endpoint Anonymity)**.** *Conditioned on $\mathcal{H}$, protocol π_{OM} achieves perfect endpoint anonymity; that is, $I(S; P \mid \mathcal{H}) = I(R; P \mid \mathcal{H}) = 0$.*

Proof. Immediate from Theorem 5. □

Endpoint Information Leakage. Theorem 5 guarantees *perfect endpoint anonymity* whenever the event $\mathcal{H}$ holds—that is, when both the first and last hops on the path are honest. If $\mathcal{H}$ does *not* hold, the adversary inevitably learns at least one endpoint. Concretely, let the sequence of adversarial hops on the path be $A_1, \ldots, A_m$ in the order they process the OM ($m \geq 1$). The earliest adversarial node A_1 can designate its immediate predecessor as the sender, and the latest adversarial node A_m can designate its immediate successor as the receiver. This limitation is intrinsic to any hop-by-hop forwarding protocol and is independent of the specific path-selection rule.

A noteworthy property of our *random-path selection algorithm* is that the probability of violating $\mathcal{H}$ depends only on the *fraction of adversarial funds*, not on the *number of adversarial nodes*. Because each hop is chosen independently (with replacement) in proportion to node funds,

$$\Pr[\neg\mathcal{H}] \;=\; 1 - \left(1 - \frac{F_A}{F_A + F_H}\right)^2, \tag{17}$$

where F_A and F_H denote the total adversarial and honest funds, respectively. Increasing the number of corrupted nodes without increasing F_A leaves the right-hand side of (17) unchanged; hence the adversary cannot amplify endpoint leakage merely by spawning additional low-fund Sybils. Whenever $\mathcal{H}$ holds, the guarantee of Corollary 2—*perfect endpoint anonymity*—remains intact.

VI. Simulations

We evaluate the analytical bounds and the proposed subset-based forwarding method using three snapshots of the public Lightning Network. Our evaluation has three objectives. First, we examine the tightness of the upper bounds in Theorems 2 and 3. Second, we compare two per-connection stress tests—the random-target and top-target adversaries—with the strong, topology- and connection-unconstrained adversary used in our analysis. Third, we determine whether the benefit of the proposed forwarding subset persists across multiple network snapshots.

A. Network Snapshots

We use snapshots of the public LN graph dated April 12, 2022, June 1, 2023, and March 19, 2026. The 2022 and 2023 snapshots were reconstructed by replaying LN gossip data obtained from the `lnresearch/topology` repository [14]. The replay procedure chronologically processes the archived gossip messages and reconstructs the public channel graph at the selected observation time.

The `lnresearch/topology` repository does not provide a public gossip archive after September 24, 2023. We therefore collected the March 19, 2026 snapshot independently. An LND node, operated through `lncli`, was connected to several high-degree LN peers and kept actively synchronized for three weeks to improve its coverage of public gossip announcements. As a macro-topological sanity check, the numbers of public nodes and channels observed by the

node were compared with the corresponding counts reported by major public LN data providers. The observed values were close to these public baselines. Nevertheless, as with any topology collected from a live node, the snapshot may be affected by incomplete gossip propagation or peer-dependent visibility.

We import the nodes and public channels in each snapshot, together with their advertised capacities, into a NetworkX graph. For a public channel $\{U_i, U_j\}$ with advertised capacity C_{ij}, the public-capacity weight of node U_i is computed as

$$F_i = \sum_{\{U_i,U_j\}\in E_L} C_{ij}. \tag{18}$$

Thus, the full advertised capacity of a public channel contributes to the public-capacity weight of each endpoint. Consequently, if

$$C_{\text{net}} = \sum_{\{U_i,U_j\}\in E_L} C_{ij}$$

denotes the aggregate capacity of the public channels, then

$$F_H = \sum_i F_i = 2C_{\text{net}}.$$

As discussed in Section II, F_i is a publicly computable policy weight. It is not an estimate of the private balance, directional liquidity, or spendable funds of U_i. In particular, we do not infer directional channel balances or impose an equal-split liquidity assumption, because OMs carry data rather than value and do not consume payment-channel liquidity.

Table II summarizes the three preprocessed snapshots and the forwarding subsets returned by Algorithm 1. The full-network public-capacity distributions are highly skewed: δ ranges from approximately 442 to 740. In contrast, the selected subsets have substantially smaller skew, with δ_S between approximately 2.1 and 4.9, where

$$\delta_S = \frac{F_{\max}^{(S)}}{F_{\text{tot}}^{(S)}/|S|}.$$

B. Simulation Setup and Adversarial Strategies

We set $l = L = 3$. Thus, an honest OM path contains two independently selected forwarding nodes and one interior directed link whose saturation can cause the OM to fail. This is the shortest setting that includes two randomized forwarding nodes. It also minimizes the failure-probability bounds in Theorems 2 and 3. We normalize the rate-limit parameter to $\alpha = 1$. This normalization does not affect the results because all relevant forwarding rates scale linearly with α.

Let W denote the forwarding-node pool: $W = V$ for the full-network experiments and $W = S$ for the subset experiments. Consistent with the communication model, any two distinct nodes in W may appear consecutively as forwarding nodes on an OM path; they need not share a public payment channel. The public channel graph G_L is used only to compute the public-capacity weights F_i and the forwarding subset.

For each of the following two per-connection adversaries, the adversary establishes an OM connection to every node $R \in W$ and sends traffic through that connection at the maximum permitted rate αF_A. Saturating one directed link $R \to T$ requires traffic at rate αF_R. Therefore, the traffic entering R can fully saturate

$$m_R = \min\left\{|W| - 1,\ \left\lfloor \frac{F_A}{F_R} \right\rfloor\right\}$$

distinct outgoing links. After one link is saturated, the remaining rate is assigned to another outgoing link, continuing until the remaining rate is less than αF_R or every possible target has been used.

Let $\mathcal{T}_R \subseteq W \setminus \{R\}$ denote the resulting set of m_R targets. For every $T \in \mathcal{T}_R$, the adversary sends a separate attack stream along a path

$$A \to R \to T \to Q_{R,T},$$

where $Q_{R,T} \in W \setminus \{R, T\}$ is the destination. Because $T \to Q_{R,T}$ is the final link to the destination, it is not subject to the forwarding rate limit. The saturated-link set is therefore

$$E_{\text{sat}} = \{(R, T) : R \in W,\ T \in \mathcal{T}_R\}.$$

Random-target adversary. For each $R \in W$, the adversary places the nodes in $W \setminus \{R\}$ in a uniformly random order and saturates links to the first m_R nodes in that order. Once these target orderings are fixed, the resulting failure probability is computed exactly over the fund-weighted path distribution of honest OMs.

Top-target adversary. For each $R \in W$, the adversary orders the nodes in $W \setminus \{R\}$ by decreasing public-capacity weight and saturates links to the first m_R nodes in that order, with ties broken deterministically. Thus, the available rate is first used to saturate the link to the highest-weight target, then the link to the second-highest-weight target, and so on. For a fixed first relay R, the contribution of its saturated outgoing links to the honest-path failure probability is proportional to

$$F_R \sum_{T\in\mathcal{T}_R} F_T.$$

The top-target ordering maximizes this quantity among all choices of m_R distinct targets for R.

The adversarial route choices determine which directed links are saturated. Honest senders continue to construct their paths using the fund-weighted random-path selection algorithm defined in Section II; thus, the failure probability is evaluated over the honest fund-weighted path distribution.

Strong optimal adversary. For comparison, we retain the topology- and connection-unconstrained adversary analyzed in Section III-A. It may saturate arbitrary directed links subject to the aggregate upper budget

$$C_A = \alpha F_A (L - 1)|W|, \tag{19}$$

and allocates this budget according to the optimal target-first ordering derived in that section. This strategy is used only as a conservative benchmark for assessing the tightness of the analytical bounds.

The random-target and top-target adversaries impose the concrete per-connection rate constraint αF_A, but still conservatively assume that the adversary can connect to every node in W and sustain its maximum permitted rate on all connections

TABLE II: Full-network and selected-subset statistics. Here, C_{net} is the aggregate capacity of the public channels, F_H is the aggregate public-capacity weight of all nodes, and S is the subset returned by Algorithm 1.

Date	**Nodes**	**Channels**	$\boldsymbol{C_{\mathrm{net}}}$ **(BTC)**	$\boldsymbol{F_H}$ **(BTC)**	$\boldsymbol{\delta}$	$\boldsymbol{\lvert S\rvert}$	$\boldsymbol{F_{\mathrm{tot}}^{(S)}}$ **(BTC)**	$\boldsymbol{F_{\mathrm{max}}^{(S)}}$ **(BTC)**	$\boldsymbol{\delta_S}$
Apr. 12, 2022	18,713	81,392	3,171.985	6,343.970	512.071	123	2,261.981	38.857	2.113
Jun. 1, 2023	15,290	64,753	3,530.386	7,060.772	442.500	119	2,494.452	45.948	2.192
Mar. 19, 2026	11,643	40,762	4,430.684	8,861.368	740.392	97	4,504.533	228.087	4.912

simultaneously. We do not impose separate CPU, bandwidth, or connection-count limits. Accordingly, these strategies are controlled per-connection stress tests rather than deployment-exact attacker models.

For $l = 3$, once the set E_{sat} of saturated directed links is determined, the failure probability can be computed exactly rather than estimated through Monte Carlo sampling. Let

$$F_H^{(W)} = \sum_{U_i \in W} F_i.$$

The exact failure probability is

$$p_{\mathrm{fail}}(W) = 1 - \left(\frac{F_H^{(W)}}{F_H^{(W)} + F_A}\right)^2 + \frac{\sum_{(U_i,U_j)\in E_{\mathrm{sat}}} F_i F_j}{\left(F_H^{(W)} + F_A\right)^2}. \tag{20}$$

The first term is the probability that at least one of the two selected forwarding nodes is adversarial. The second term is the probability that both forwarding nodes are honest and the directed link between them has been saturated.

C. Common BTC Benchmark

To compare the snapshots without introducing variation from historical BTC–USD exchange rates, we express F_A in BTC and use the same BTC attack weight in every snapshot. Using the BTC/USD reference price of USD 69,939.88 per BTC on March 19, 2026 [15], we define

$$B = \frac{\text{USD } 1{,}000{,}000}{\text{USD } 69{,}939.88/\text{BTC}} \approx 14.30 \text{ BTC}. \tag{21}$$

Thus, B was equivalent to approximately USD 1 million at the reference exchange rate, but the same B BTC is applied to all three snapshots.

At $F_A = B$, the adversarial public-capacity fractions

$$\gamma = \frac{F_A}{F_A + F_H}$$

are approximately 0.225%, 0.202%, and 0.161% for the 2022, 2023, and 2026 snapshots, respectively.

D. Full-Network Results

Figure 1 shows the failure probability when all public nodes are available for forwarding-node selection. Across all three snapshots, the top-target and strong-optimal curves coincide. Both also closely track the upper bound of Theorem 2.

At the common benchmark $F_A = B$, the top-target and strong optimal adversaries cause failure probabilities of 45.953%, 40.919%, and 56.720% in the 2022, 2023, and 2026 snapshots, respectively. The corresponding analytical bounds are 46.485%, 41.665%, and 58.691%. Thus, the observed failure probability is within approximately two percentage points of the analytical bound in every snapshot.

The random-target adversary is substantially less damaging. At $F_A = B$, it causes failure probabilities of only 0.908%, 0.779%, and 0.555% in the three snapshots, respectively. The large difference between the random-target and top-target strategies demonstrates the importance of directing the available traffic toward high-weight targets.

The 2026 snapshot has the smallest adversarial fraction γ at the common benchmark but the largest failure probability under the top-target and strong-optimal strategies. This behavior is consistent with its greater fund skew, $\delta = 740.392$, and illustrates that resilience depends not only on the adversarial fraction but also on the distribution of the honest nodes' public-capacity weights.

E. Subset-Based Forwarding Results

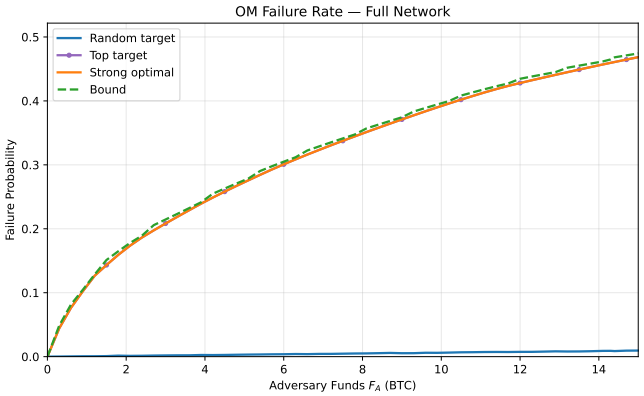
We next restrict forwarding-node selection to the subset returned by Algorithm 1. The selected subsets contain 123, 119, and 97 nodes in the 2022, 2023, and 2026 snapshots, respectively.

Figures 2–4 show the results using four-panel figures. In each figure, the lower-right panel corresponds to the subset returned by Algorithm 1. The other panels show comparison subsets constrained to have the cardinality indicated in the panel title.

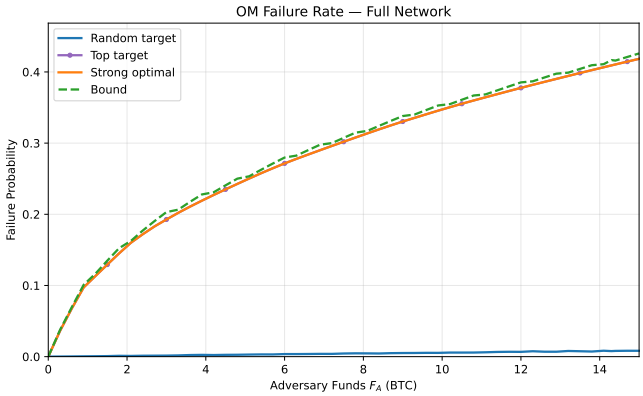
Restricting forwarding to the selected subset substantially reduces the failure probability under the top-target and strong-optimal strategies. Under the strong optimal adversary, the failure probability falls from 45.953%, 40.919%, and 56.720% on the full networks to 3.888%, 3.602%, and 3.730% on the selected subsets. These correspond to reductions by factors of approximately 11.8, 11.4, and 15.2, respectively.

At the common benchmark, the top-target adversary produces the same values as the strong optimal adversary in every snapshot. Their curves are also nearly indistinguishable throughout the selected-subset experiments. Thus, although the top-target adversary is subject to the per-connection rate constraint, its target ordering makes it nearly as effective as the topology- and connection-unconstrained benchmark.

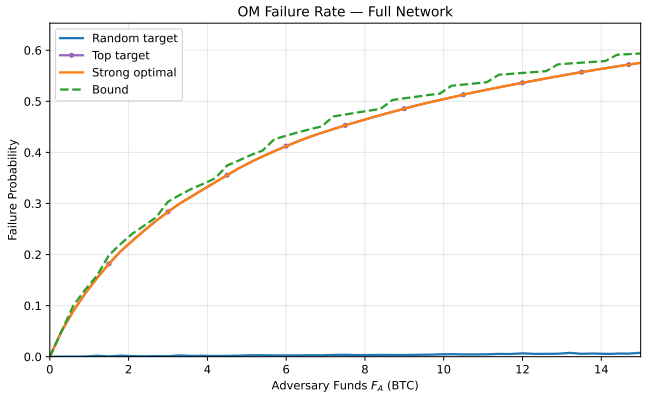
The random-target adversary behaves differently. Its failure probabilities are 2.517%, 2.340%, and 1.216% on the selected subsets, compared with 0.908%, 0.779%, and 0.555% on



(a) April 12, 2022.



(b) June 1, 2023.



(c) March 19, 2026.

Fig. 1: OM failure probability when forwarding nodes are selected from the full network. For each first relay, the random-target adversary orders the possible targets uniformly at random, whereas the top-target adversary orders them by decreasing public-capacity weight. Each adversary saturates as many outgoing links as the rate available through that relay permits. The strong optimal adversary is the topology- and connection-unconstrained aggregate-budget benchmark. The horizontal axis is the adversarial public-capacity weight F_A in BTC.

TABLE III: OM failure probabilities at $F_A = B \approx 14.30$ BTC. All entries are percentages. "Bound" refers to Theorem 2 for the full network and Theorem 3 for the selected subset.

Date	**Forwarding pool**	**Random-target**	**Top-target**	**Strong optimal**	**Bound**
Apr. 12, 2022	Full network	0.908	45.953	45.953	46.485
Apr. 12, 2022	Selected subset ($\lvert S\rvert = 123$)	2.517	3.888	3.888	6.607
Jun. 1, 2023	Full network	0.779	40.919	40.919	41.665
Jun. 1, 2023	Selected subset ($\lvert S\rvert = 119$)	2.340	3.602	3.602	6.172
Mar. 19, 2026	Full network	0.555	56.720	56.720	58.691
Mar. 19, 2026	Selected subset ($\lvert S\rvert = 97$)	1.216	3.730	3.730	6.871

the full networks. Thus, subset restriction does not reduce the failure probability under this dispersed attack strategy, although its absolute failure probability remains below that of the top-target and strong-optimal adversaries.

The bounds from Theorem 3, equal to 6.607%, 6.172%, and 6.871%, remain above all three evaluated strategies.

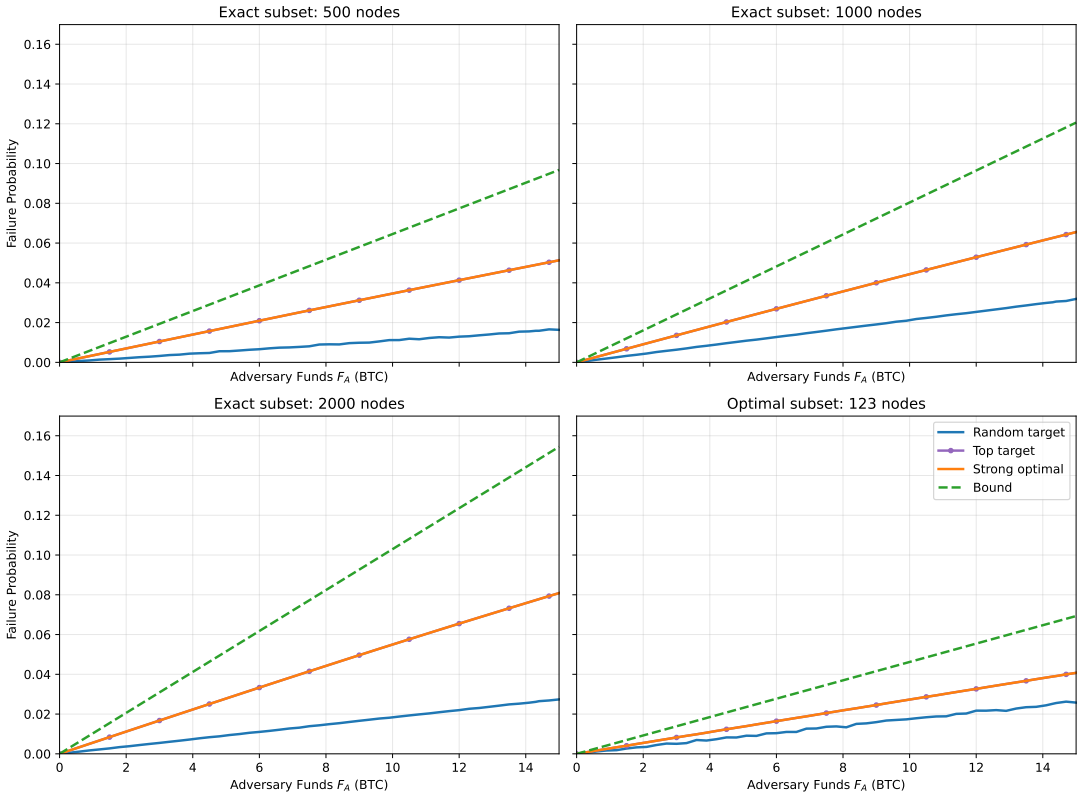


Fig. 2: Subset results for the April 12, 2022 snapshot. The lower-right panel shows the 123-node subset returned by Algorithm 1; the other panels show fixed-cardinality comparison subsets.

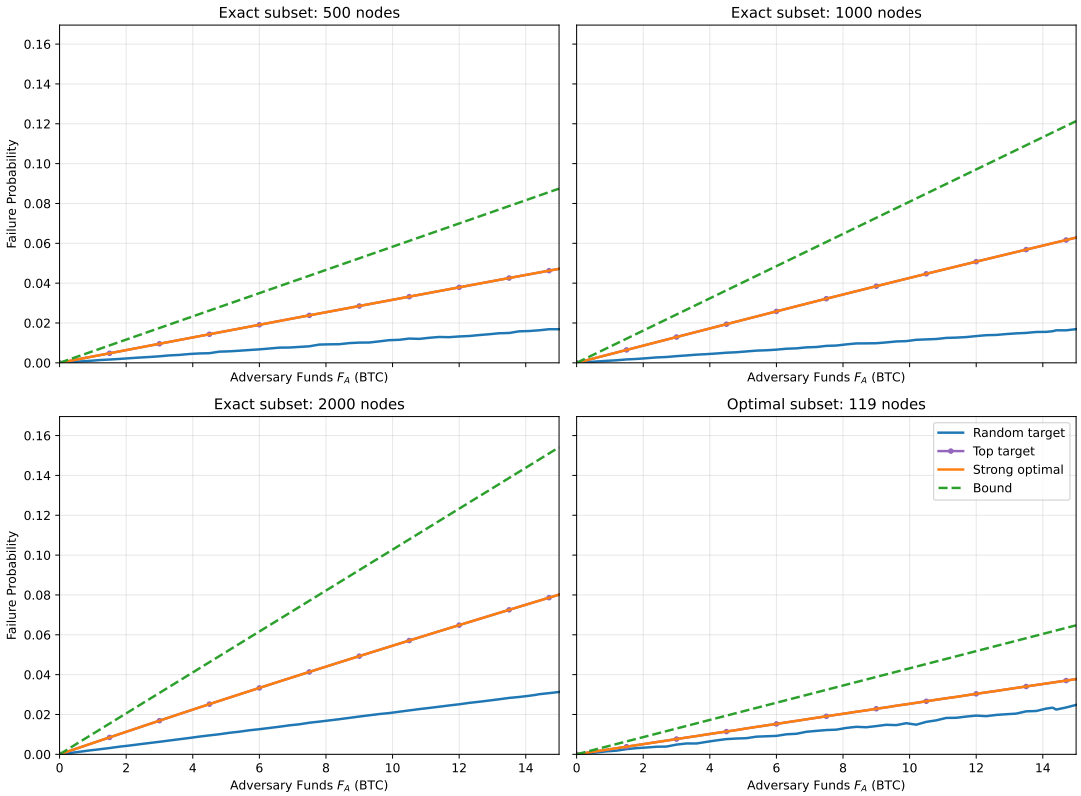


Fig. 3: Subset results for the June 1, 2023 snapshot. The lower-right panel shows the 119-node subset returned by Algorithm 1; the other panels show fixed-cardinality comparison subsets.

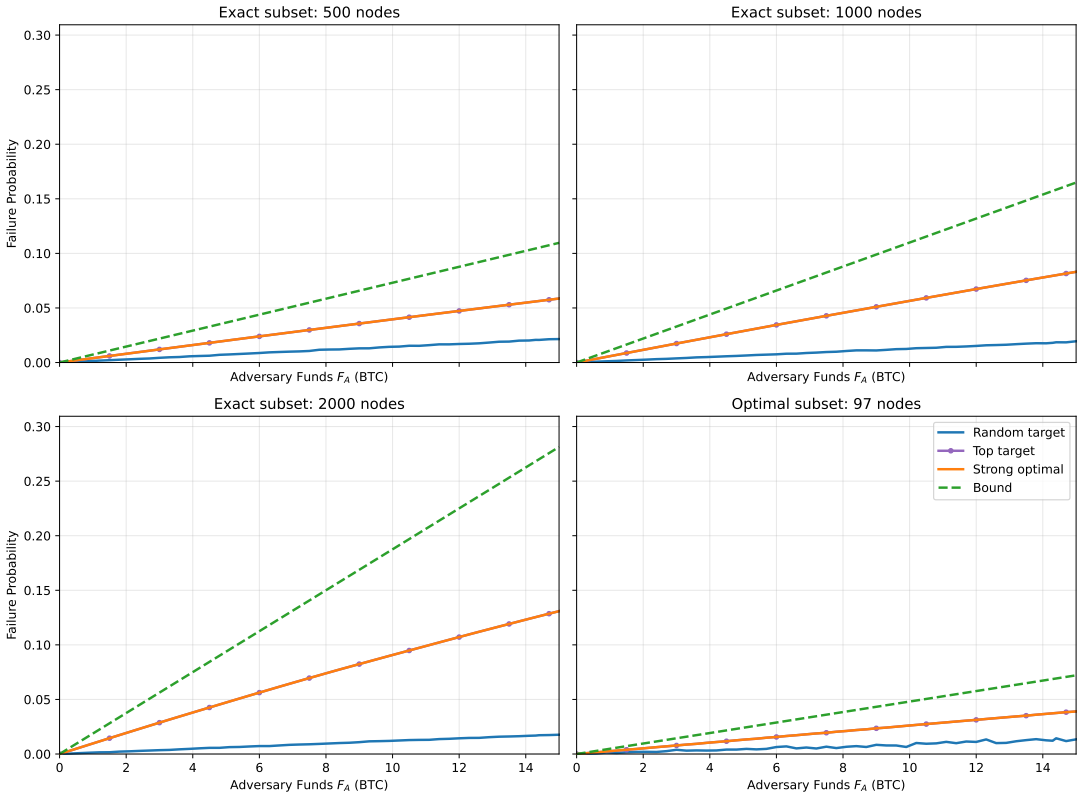


Fig. 4: Subset results for the March 19, 2026 snapshot. The lower-right panel shows the 97-node subset returned by Algorithm 1; the other panels show fixed-cardinality comparison subsets.

Forwarding concentration. Some concentration is already present under full-network routing because forwarding nodes are selected in proportion to their public-capacity weights. Let Λ denote the aggregate rate of honest OMs, let W denote the forwarding pool, and define

$$F_{\text{tot}}^{(W)} = \sum_{U_j \in W} F_j.$$

Since each OM selects $l-1$ forwarding nodes, the expected forwarding load at $U_i \in W$ is

$$\lambda_i^{(W)} = (l-1)\Lambda \frac{F_i}{F_{\text{tot}}^{(W)}}.$$

Because the network snapshots do not contain network-wide honest-OM traffic measurements, we express the load as a function of Λ. Under full-network weighted routing, the nodes subsequently selected into S account for 35.7%, 35.3%, and 50.8% of forwarding positions in the 2022, 2023, and 2026 snapshots, respectively. Restricting forwarding to S therefore increases every selected relay's expected load by $\frac{F_H}{F_{\text{tot}}^{(S)}}$, which equals 2.80, 2.83, and 1.97, respectively. Thus, the incremental burden caused by subset restriction is approximately a two- to three-fold increase relative to the full-network weighted baseline.

The largest selected relay accounts for 1.72%, 1.84%, and 5.06% of forwarding positions in the three snapshots. For $l = 3$, if that relay becomes unavailable while remaining in a sender's local view, the probability that a selected path contains it is 3.41%, 3.65%, and 9.87%, respectively. The fixed-cardinality subsets evaluated in Figures 2–4 provide larger-pool alternatives for deployments that impose a minimum cardinality. Designing rotating-subset policies is outside the scope of the present work.

Finally, OMs are designed as a lightweight, unreliable communication mechanism and may be retransmitted over independently selected paths. The largest selected-subset failure probability observed at the common benchmark is 3.888%. Two independent attempts therefore both fail with probability

$$(3.888\%)^2 \approx 0.151\%.$$

The corresponding probability of successful delivery is approximately 99.85%.

Overall, the results are consistent across the three snapshots. The top-target adversary is substantially more effective than the random-target adversary and, at the common benchmark, matches the strong optimal adversary. Nevertheless, restricting forwarding to the subset selected by Algorithm 1 reduces the top-target and strong-optimal failure probabilities from 40.9%–56.7% to 3.6%–3.9%. The random-target failure probability remains between approximately 1.2% and 2.5% on the selected subsets.

Table III compares the full-network and selected-subset results at the common benchmark $F_A = B$.

VII. Conclusion

In this work, we analyzed the resilience of onion messaging against DoS attacks under a policy that uses publicly observable channel capacities to set forwarding limits and relay-

selection probabilities. We showed that this policy provides Sybil resistance because an adversary's influence depends on its aggregate public-capacity weight, rather than on the number of identities it creates. However, its resilience can deteriorate when public-capacity weights are highly skewed, as they are in LN. To address this limitation, we proposed restricting forwarding-node selection to a carefully selected subset of nodes and provided a quadratic-time algorithm for identifying a subset that minimizes the derived failure-probability bound.

For the evaluated setting $l = L = 3$, our evaluation over three LN snapshots shows that the proposed method substantially improves resilience under both the top-target per-connection adversary and the strong optimal aggregate-budget adversary. At the common benchmark $F_A \approx 14.30$ BTC—approximately USD 1 million at the reference exchange rate used in our evaluation—the two adversaries produce the same failure probability in each snapshot: between 40.9% and 56.7% when the full network is available for forwarding-node selection, and between 3.6% and 3.9% when selection is restricted to the optimized subsets. The corresponding subset-based analytical upper bounds range from 6.2% to 6.9%. The random-target adversary is substantially weaker, producing failure probabilities of 0.6%–0.9% over the full networks and 1.2%–2.5% over the selected subsets. Thus, subset restriction does not improve resilience against this dispersed strategy, although its absolute failure probability remains below that of the targeted strategies.

These empirical conclusions are limited to $l = L = 3$. Although the analytical bounds characterize the effect of general path parameters, the optimized subsets and the performance of the evaluated per-connection adversaries may change for longer paths.

Because OMs are designed as a lightweight, unreliable communication mechanism, additional reliability can be obtained through independent retransmissions. Based on the largest selected-subset failure probability of 3.888% at the benchmark, two independent attempts deliver an OM with probability approximately 99.85%.

Future work may study empirical sensitivity to longer paths, including the resulting optimized subsets and per-connection attack performance, as well as dynamic subset recomputation, rotating forwarding subsets, and adaptive per-peer throttling.

REFERENCES

[1] J. Poon and T. Dryja, "The bitcoin lightning network: Scalable off-chain instant payments," https://lightning.network/lightning-network-paper.pdf, 2016.

[2] C. Decker and R. Wattenhofer, "A fast and scalable payment network with bitcoin duplex micropayment channels," in *Stabilization, Safety, and Security of Distributed Systems: 17th International Symposium, SSS*. Springer, 2015, pp. 3–18.

[3] P. McCorry, M. Möser, S. F. Shahandashti, and F. Hao, "Towards bitcoin payment networks," in *Information Security and Privacy: 21st Australasian Conference, ACISP 2016*. Springer, 2016, pp. 57–76.

[4] C. L. Team, "Announcing c-lightning v0.9.3—"federal qualitative strengthening"," January 2021, blockstream Blog. Accessed: 2025-05-08. [Online]. Available: https://blog.blockstream.com/announcing-c-lightning-v0-9-3-federal-qualitative-strengthening/

[5] V. Wallace, "Onion messages demystified," 2022, lightning Dev Kit Blog. Accessed: 2025-05-08. [Online]. Available: https://lightningdevkit.org/blog/onion-messages-demystified

[6] L. Developers, "LNURL documents," 2024, gitHub Repository. Accessed: 2025-05-08. [Online]. Available: https://github.com/lnurl/luds

[7] L. N. Developers, "Bolt 4: Onion messages," 2024, gitHub Repository. Accessed: 2025-05-08. [Online]. Available: https://github.com/lightning/bolts/blob/master/04-onion-routing.md#onion-messages

[8] D. Goldschlag, M. Reed, and P. Syverson, "Onion routing," *Communications of the ACM*, vol. 42, no. 2, pp. 39–41, 1999.

[9] C. Shikhelman and S. Tikhomirov, "Unjamming lightning: A systematic approach," *Cryptology ePrint Archive*, vol. 2022, p. Report 1454, 2022. [Online]. Available: https://eprint.iacr.org/2022/1454

[10] A. Bashiri and M. Khabbazian, "Short paper: Onion messages on leash," in *International Conference on Financial Cryptography and Data Security (FC) 2024*. Springer, 2024, pp. 283–292.

[11] t-bast, "Onion Messages Rate-Limiting," https://gist.github.com/t-bast/e37ee9249d9825e51d260335c94f0fcf, gitHub Gist, last updated October 22, 2024; accessed July 28, 2026.

[12] N. Borisov, G. Danezis, P. Mittal, and P. Tabriz, "Denial of service or denial of security? How attacks on reliability can compromise anonymity," in *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS'07)*. New York, NY, USA: Association for Computing Machinery, 2007, pp. 92–102.

[13] L. Zhuang, F. Zhou, B. Y. Zhao, and A. Rowstron, "Cashmere: Resilient anonymous routing," in *Proceedings of the 2nd Symposium on Networked Systems Design and Implementation (NSDI'05)*. Boston, MA, USA: USENIX Association, 2005, pp. 301–314. [Online]. Available: https://www.usenix.org/legacy/event/nsdi05/tech/full_papers/zhuang/zhuang.pdf

[14] C. Decker, "Lightning network research—topology datasets," https://github.com/lnresearch/topology, accessed: July 27, 2026.

[15] StatMuse Money, "BTC March 19th," https://www.statmuse.com/money/ask/btc-march-19th, 2026, accessed: July 27, 2026; reference price reported for March 19, 2026: USD 69,939.88 per BTC.