APPENDIX A
INTERVAL SYNCHRONIZATION

In our secure sketch PQM protocol (Section III-B), the 'Interval End' and 'Report' control messages can be used to synchronize the interval number between Alice and Bob, even if the path between them is subject to variable latency. However, even in the benign case, out-of-order packet delivery at the network layer can cause packets in an interval u to arrive after the 'Interval End' message u (and thus be interpreted by Bob as part of interval $u+1$). Note that out-of-order packet delivery could also occur if Eve deliberately delays packets. To avoid false alarms due to more than αT packets arriving out-of-order before the 'Interval End' control message, we can tune parameter α.

To ensure natural packet reordering does not cause a loss of interval synchronization between the sender and receiver, a good rule of thumb is to ensure that $\alpha T \geq 99^{th}$ percentile of packet lag. Define the packet lag as the number of packets that were sent by the sender after a reordered packet, but were received at the receiver earlier than the reordered packet itself (*e.g.,* if a sender sends the stream 1,2,3,4,5,6,7,8 but the received stream is 1,2,4,5,6,7,3,8, then packet 3 is the reordered packet and packet lag is 4). The value of the packet lag depends on the the class of packets monitored by the PQM protocol. If the packets belong to the same network flow, we can safely assume that packet lag is less than 128 packets, because this is the assumption made in IPSec. Thus, it suffices to take $\alpha T > 1280$. In cases when multiple network flows are monitored with the same PQM instance, then packet lag can be very high (due to load balancing, ECMP, *etc.*); however, we conjecture that even if there is a 10ms difference between the "fast path" used by one group of flows and the "slow path" used by another group of flows, for 1 Gbps flow of traffic, packet lag should be on the order of 10^9 bps/64 bytes/packet $\times 0.01$ sec $= 1.6 \times 10^5$ packets, so we can use $\alpha T > 1.6 \times 10^6$.

APPENDIX B
FAST PACKET HASHING

Section III-F indicated that the computational cost of packet hashing can be reduced by (1) first mapping packets from from U to a short n_1-bit string using an efficient ε_g-almost universal hash function, and (2) then using a PRF or 4-wise independent hash to map from this n_1-bit string to the sketch. We show this is possible via approaches based on [47].

Preliminaries. Return to the notation of Section III-C, and recall that U is the universe of all possible packets, $\mathbf{v}$ is the characteristic vector of the stream of packets, and $\mathbf{w}$ is the sketch vector of length N. Let $g : U \to \{0,1\}^{n_1}$ be an ε_g-almost universal hash function, as defined in Section III-F. The hash function g maps the packet stream containing elements in U to a new 'intermediate' stream where each element is an n_1-bit string. Let $\mathbf{u}$ be an 'intermediate vector' which is the characteristic vector of this new stream of n_1-bit strings.

Our approach amounts to using the ε_g-almost universal hash g to hash $\mathbf{v}$ the 'intermediate vector' $\mathbf{u}$, and then using a a second-moment estimation scheme to hash $\mathbf{u}$ down to the sketch $\mathbf{w}$. Thus, the second-moment estimation scheme estimates the second moment of $\mathbf{u}$, rather than the real characteristic vector $\mathbf{v}$! We now show that, if ε_g is sufficiently small, this does very little damage, since $\|\mathbf{u}\|_2 \approx \|\mathbf{v}\|_2$.

Theorem B.1. *Given a vector $\mathbf{v} \in 2^{|U|}$ and $\mathbf{u} \in \mathbb{R}^{2^{n_1}}$. Then if $g : U \to \{0,1\}^n$ is an ε_g-almost 2-wise independent hash function per equation (14), is used to map $\mathbf{v}$ to $\mathbf{u}$ according to the algorithm $u_{g(x)} += v_x$ (*i.e., *$\forall\, x \in \mathbf{v}$ the $g(x)^{th}$ counter in $\mathbf{u}$ is incremented with value v_x) then*

$$\Pr\left[\, \|\mathbf{u}\|_2 - \|\mathbf{v}\|_2 \, | > \delta_1 \|\mathbf{v}\|_2\right] < \delta_2 \tag{18}$$

as long as $|\mathbf{v}|_1 > \frac{\delta_1 \delta_2}{\varepsilon_g}$.

To apply this theorem, recall from Section III-C that $|\mathbf{v}|_1 = A + D$. Thus, for (α, β, δ)-secure PQM we would like (18) to hold when $D = \alpha T$ and $D = \beta T$, with $\delta_1 \ll \varepsilon = \frac{\beta - \alpha}{\alpha + \beta}$. We will conservatively take $|\mathbf{v}|_1 = T$, and $\delta_1 = \frac{\varepsilon}{10}$ and set $\delta_2 = \frac{\delta}{100}$. Then (α, β, δ)-secure PQM require the hash function g to have ε_g as in (15) because $\varepsilon_g < \frac{\varepsilon\delta}{10^3 T} = \frac{\delta}{10^3 T}\frac{\beta - \alpha}{\alpha + \beta}$.

Proof of Theorem B.1. Let v_a be the a^{th} entry of characteristic vector $\mathbf{v}$. Now, start with the observation that

$$\begin{aligned} \|\mathbf{u}\|_2^2 &= \sum_{g(a)=g(b)} v_a v_b \\ &= \sum_a v_a^2 + \sum_{a \neq b, g(a)=g(b)} v_a v_b \\ &= \|\mathbf{v}\|_2^2 + \sum_{a \neq b} v_a v_b Y_{a,b} \end{aligned} \tag{19}$$

where we define the random variable $Y_{a,b}$ as

$$Y_{a,b} = \begin{cases} 1 & \text{if } g(a) = g(b), a \neq b, \\ 0 & \text{else.} \end{cases}$$

and from (19) we take the expectation over the randomness in g and find that

$$\begin{aligned} \mathbb{E}[\, | \, \|\mathbf{u}\|_2^2 - \|\mathbf{v}\|_2^2 \, | \,] &\leq \sum_{a,b} |v_a v_b| \mathbb{E}[|Y_{a,b}|] \\ &\leq \sum_{a,b} |v_a v_b| \cdot \varepsilon_g \\ &= (|\mathbf{v}|_1^2 - \|\mathbf{v}\|_2^2) \cdot \varepsilon_g \end{aligned} \tag{20}$$

where the first inequality follows from (19), the second inequality follows because per equation (14) the collision probability of g is ε_g.

Now, we would like to ensure that $\|\mathbf{u}\|_2$ provides a good estimate of $\|\mathbf{v}\|_2$. That is, we would like to satisfy (18). Using Markov's inequality, we have

$$\begin{aligned} \Pr\left[\left|\|\mathbf{u}\|_2^2 - \|\mathbf{v}\|_2^2\right| > \delta_1 \|\mathbf{v}\|_2^2\right] &\leq \frac{\mathbb{E}[\, | \, \|\mathbf{u}\|_2^2 - \|\mathbf{v}\|_2^2 \, | \,]}{\delta \|\mathbf{v}^2\|_2} \\ &\leq \frac{(|\mathbf{v}|_1^2 - \|\mathbf{v}\|_2^2)}{\|\mathbf{v}\|_2^2} \frac{\varepsilon_g}{\delta_1} \qquad \text{(From (20))} \\ &\leq |\mathbf{v}|_1 \frac{\varepsilon_g}{\delta_1} \end{aligned}$$

And rearranging the last inequality we know that (18) holds as long as $|\mathbf{v}|_1 > \frac{\delta_1 \delta_2}{\varepsilon_g}$ which completes the proof. □

APPENDIX C
PROOF OF THEOREM V.1

First, the probability that any efficient adversary Eve successfully forges the interval end message or onion report of an honest node (by forging the MAC) is negligible. We argue that Eve does not tamper with the control messages:

Claim C.1. *If Eve tampers with the 'Interval End' message or any θ_i in the 'Onion Report' message, then Alice will localize a node adjacent to Eve.*

Proof. Let R_E be the upstream-most node where Eve tampered with either the 'Interval End' message or the 'Onion Report'. Let R_j be the first honest node that is downstream of node R_E (we know such a node exists because Eve cannot occupy Bob's node). Since all the $R_1, ..., R_{E-1}$ behave honestly, their all reports $\theta_1, ..., \theta_{E-1}$ will be present and valid. Also, conditioned on Eve not forging R_j's MAC, θ_j will either be invalid (*e.g.,* if Eve tampered with some θ_ℓ for $\ell > j$, since θ_ℓ is nested inside θ_j) or missing (*e.g.,* if Eve dropped the 'Interval End' message). It follows that the upstream-most invalid report θ_x occurs on some link between R_{E-1} and R_j, so that Alice will output a link adjacent to Eve. □

We may now suppose that Alice receives correct reports from all honest nodes. We next present some notation.

Notation. Let D_i be a count of the number of failures that occurred on the path between Alice and R_i. Let $\mathbf{v}_A$ be the characteristic vector of the stream of packets that Alice sends and let $\mathbf{v}_i$ for $i \in [K+1]$ be the characteristic vector of the stream of data packets that R_i receives. Let $\mathbf{x}_i = \mathbf{v}_A - \mathbf{v}_i$. As in equation (3), we can decompose $\mathbf{x}_i$ into two vectors $\mathbf{x}_i = \mathbf{d}_i + \mathbf{a}_i$, where $\mathbf{d}_i$ is the characteristic vector of packets *dropped* on the path from Alice to R_i, and contains the non-negative components of $\mathbf{x}_i$. The vector $\mathbf{a}$ is the characteristic vector of packets *added* on the path from Alice to R_i, and contains the non-positive components of $\mathbf{x}_i$.

The following lemma, proved in Appendix C-A of [25], proves the "few false positives" and "secure localization" conditions of Definition IV.1:

Lemma C.2. *Let $\Gamma = \frac{T}{K+1}\frac{\beta(2\alpha+\beta)}{\alpha+2\beta}$ and $\varepsilon_i = \frac{1}{2i}\frac{\beta-\alpha}{2\beta+\alpha}$. For every $i \in [K]$, assume that R_i computes an estimate V_i that (ε_i, δ')-estimates $\|\mathbf{x}_i\|_2^2$. Suppose also that $\|\mathbf{x}_i\|_2^2 \leq \frac{\beta i}{K+1}$. Then with probability at least $1-2\delta'$ it follows that:*

1) *If "link $(i, i+1)$ is good" so that $\|\mathbf{x}_{i+1}\|_2^2 - \|\mathbf{x}_i\|_2^2 \leq \frac{\alpha}{K+1}T$ then $V_{i+1} - V_i \leq \Gamma$.*
2) *If "link $(i, i+1)$ is bad" so that $\|\mathbf{x}_{i+1}\|_2^2 - \|\mathbf{x}_i\|_2^2 \geq \frac{\beta}{K+1}T$ then $V_{i+1} - V_i \geq \Gamma$.*

Few false positives: To prove this, we consider an interval where all the nodes on the path behave honestly. During this interval, we know that no packets were added anywhere on the path (so that $\|\mathbf{a}_i\|_2^2 = 0$ for each $i \in [K+1]$) and less than $\frac{\alpha}{K+1}$ packets were dropped at each link. We can apply equation (3) to find that for each link $(i, i+1)$ we have

$$\|\mathbf{x}_{i+1}\|_2^2 - \|\mathbf{x}_i\|_2^2 = (D_{i+1} + 0) - (D_i + 0) \leq \frac{\alpha}{K+1} \quad (21)$$

and the telescoping nature of (21) gives us that

$$\|\mathbf{x}_i\|_2^2 = (\|\mathbf{x}_i\|_2^2 - \|\mathbf{x}_{i-1}\|_2^2) + ... + (\|\mathbf{x}_2\|_2^2 - \|\mathbf{x}_1\|_2^2) + \|\mathbf{x}_1\|_2^2 \leq \frac{\alpha i}{K+1} \quad (22)$$

We can now apply Lemma C.2 to show that, with probability at least $1-2\delta'$ we have that $V_{i+1} - V_i \leq \Gamma$ so that Alice will not output link $(i, i+1)$. A union bound over the $K+1$ links gives us that Alice will output $\surd$ during this interval with probability at least $1 - 2(K+1)\delta'$.

Secure localization: We now show that if Eve causes more than a β fraction of failures in the interval, then with probability at least $1-\delta$, Alice will either catch Eve or output a link with more than $\frac{\alpha}{K+1}$ failures. Recall that Alice outputs the upstream-most link $\ell = (i, i+1)$ for which there is an "alarm", *i.e.,* where $V_{i+1} - V_i \geq \Gamma$. We need the following simple observation:

Lemma C.3. *Define event E_i as the event that $\|\mathbf{x}_i\|_2^2 \leq \frac{\beta i}{K+1}$ .For each $i \in [K+1]$, if Alice does not raise an alarm for any link upstream of link i, then E_i holds with probability $1-2i\delta'$.*

Proof. Suppose that Alice does *not* raise an alarm for all links upstream of node R_i. Lemma C.2 implies that $\|\mathbf{x}_{j+1}\|_2^2 - \|\mathbf{x}_j\|_2^2 \leq \frac{\beta}{K+1}$ with probability $1-2\delta'$, for each link $(j, j+1)$ where $j \in [i-1]$. The lemma follows from a union bound over these links and a telescoping sum as in (22). □

First we show that the with high probability Alice will not output an honest link. Let link $(i, i+1)$ be "honest", *i.e.,* have a fewer than $\frac{\alpha}{K+1}$ failures, and assume that Alice does not raise alarm for any links upstream of R_i. Now, Lemma C.2 shows that, conditioned on E_i, Alice will not raise an alarm for link $(i, i+1)$ with probability at least $1-2\delta'$. Since Alice does not alarm for any links upstream of R_i, we can apply Lemma C.3 to remove the conditioning on E_i. It follows that Alice will not output honest link $(i, i+1)$ with probability at least $1 - 2(i+1)\delta'$. Taking a union bound over all honest links gives that Alice will not alarm for any honest link with probability at least $1 - 2(K+1)^2\delta'$. Next, we need to show that Alice either will raise an alarm for a link adjacent to Eve or link with more than $\frac{\alpha}{K+1}$ failures. The proof hinges on the following technical lemma, proved in Appendix C-B of our technical report [25]:

Lemma C.4. *If Eve occupies $M \leq \sqrt{(K+1)(1-\frac{\rho}{\beta}K^2)}$ links and causes a β-fraction of failures in the interval, then there must be a link $(i, i+1)$ that is adjacent to Eve with*

$$\|\mathbf{x}_{i+1}\|_2^2 - \|\mathbf{x}_i\|_2^2 \geq \frac{\beta}{K+1}T \quad (23)$$

Now let link $(i, i+1)$ be the upstream-most link that is adjacent to Eve such that (23) holds. (Lemma C.4 guarantees the existence of such a link.) We have two cases:

1. Suppose Alice did not raise an alarm for a link upstream of R_i. Combining Lemma C.3 and Lemma C.2 it follows that Alice will alarm for link $(i, i+1)$ adjacent to Eve with probability $1 - 2(i+1)\delta'$.

2. Suppose Alice did raise an alarm for a link upstream of R_i. It follows from Lemma C.2 that there is some link $(j, j+1)$

for $j \leq [i-1]$ where, with probability $1-2\delta'$,

$$\tfrac{\alpha}{K+1} \leq \|\mathbf{x}_{j+1}\|_2^2 - \|\mathbf{x}_j\|_2^2 = D_{j+1} - D_j + \|\mathbf{a}_{j+1}\|_2^2 - \|\mathbf{a}_j\|_2^2$$

where the equality comes from applying equation (3). Now if link $(j, j+1)$ is adjacent to Eve, it follows that Alice alarms for a link adjacent to Eve, and we are done. Thus, suppose that link $(j, j+1)$ is *not* adjacent to Eve. Then, it follows that no new packets could have been added to this link, and so we have that $\|\mathbf{a}_{j+1}\|_2^2 = \|\mathbf{a}_j\|_2^2$. Thus, if link $(j, j+1)$ is *not* adjacent to Eve, then Alice must have raised an alarm for a link with $D_{j+1} - D_j \geq \frac{\alpha}{K+1}$ failures, as required.

Combining these cases, we see that with probability at least $1 - 2(K+1)\delta'$, Alice will either raise an alarm for a link that is either (a) adjacent to Eve, or (b) has more than $\frac{\alpha}{K+1}$ failures, as required.

Sizing the sketches. To ensure that (α, β, δ)-statistical security holds, we take $\delta' = \delta/4(K+1)^2$. Next, recall that Lemma C.2 requires sketches that (ε_i, δ')-estimate the p^{th} moment with $\varepsilon_i = \frac{1}{2i}\frac{\beta-\alpha}{2\beta+\alpha}$. For simplicity, we now suppose that the sketches are constructed using 4-wise independent hashing functions, so we plug ε_i, δ in Theorem III.1 to find that for $i \in [K+1]$ it suffices to take sketches $\mathbf{w}_i$, $\mathbf{w}_i^A$ of with $N_i > \frac{2}{\varepsilon_i^2 \delta}$, where the number of bits per counter is as in (6). Substituting in the values for ε_i, δ' gives us (17) as required.

A. Proof of Lemma C.2

From the statement of the lemma, we have that R_i computes an estimate V_i that (ε_i, δ')-estimates $\|\mathbf{x}_i\|_2^2$ for every $i \in [K]$. That is:

$$\Pr\left[\left|V_i - \|\mathbf{x}_i\|_2^2\right| \leq \varepsilon_i \|\mathbf{x}_i\|_2^2\right] < 1 - \delta' \tag{24}$$

We now prove each item separately.

Link $(i, i+1)$ is good. Since V_i $(\varepsilon_i, \delta_i)$-approximates $\|\mathbf{x}_i\|_2^2$, we can apply (24) to find, that with probability $1-2\delta'$,

$$\begin{aligned}
V_{i+1} - V_i &\leq (1+\varepsilon_{i+1})\|\mathbf{x}_{i+1}\|_2^2 + (1-\varepsilon_i)\|\mathbf{x}_i\|_2^2 \\
&\leq (1+\varepsilon_{i+1})(\|\mathbf{x}_{i+1}\|_2^2 - \|\mathbf{x}_i\|_2^2) + (\varepsilon_{i+1}+\varepsilon_i)\|\mathbf{x}_i\|_2^2 \\
&\leq (1+\varepsilon_{i+1})\tfrac{\alpha}{K+1}T + (\varepsilon_{i+1}+\varepsilon_i)\tfrac{i\beta}{K+1}T \\
&= \tfrac{\alpha}{K+1}T\left(1 + \varepsilon_{i+1}(1+\tfrac{\beta}{\alpha}i) + \varepsilon_i i\tfrac{\beta}{\alpha}\right) \\
&\leq \tfrac{\alpha}{K+1}T\left(1 + (i+1)\varepsilon_{i+1}(1+\tfrac{\beta}{\alpha}) + i\varepsilon_i(1+\tfrac{\beta}{\alpha})\right) \\
&= \tfrac{T}{K+1}\tfrac{\beta(2\alpha+\beta)}{\alpha+2\beta} = \Gamma
\end{aligned} \tag{25}$$

where we get the required inequality by putting $\varepsilon_i = \frac{1}{2i}\frac{\beta-\alpha}{2\beta+\alpha}$.

Link $(i, i+1)$ is bad. Again, we apply (24) to find, that with probability $1-2\delta'$,

$$\begin{aligned}
V_{i+1} - V_i &\geq (1-\varepsilon_{i+1})(\|\mathbf{x}_{i+1}\|_2^2 - \|\mathbf{x}_i\|_2^2) - (\varepsilon_{i+1}+\varepsilon_i)\|\mathbf{x}_i\|_2^2 \\
&\geq (1-\varepsilon_{i+1})\tfrac{\beta}{K+1}T - (\varepsilon_{i+1}+\varepsilon_i)\tfrac{i\beta}{K+1}T \\
&= \tfrac{T}{K+1}\tfrac{\beta(2\alpha+\beta)}{\alpha+2\beta} = \Gamma
\end{aligned} \tag{26}$$

where we again get the required inequality by putting $\varepsilon_i = \frac{1}{2i}\frac{\beta-\alpha}{2\beta+\alpha}$.

B. Proof of Lemma C.4

Since Eve occupies M links and causes at least a β-fraction failures, it immediately follows that there exists a link $(i, i+1)$ adjacent to Eve where at least $\frac{\beta}{M}$-fraction of failures, *i.e.,* $D_{i+1} - D_i \geq \frac{\beta}{M}$. Now if the following holds

$$\|\mathbf{x}_{i+1}\|_2^2 - \|\mathbf{x}_i\|_2^2 > \tfrac{\beta}{K+1}T \tag{27}$$

we are done, since link $(i, i+1)$ is adjacent to Eve. Thus, suppose (27) do *not* hold. Then, applying identity (3), we have that

$$\begin{aligned}
\tfrac{\beta}{K+1}T &\geq \|\mathbf{x}_{i+1}\|_2^2 - \|\mathbf{x}_i\|_2^2 \\
&= D_{i+1} - D_i + \|\mathbf{a}_{i+1}\|_2^2 - \|\mathbf{a}_i\|_2^2
\end{aligned}$$

rearranging and then using that fact that $D_{i+1} - D_i \geq \frac{\beta}{M}$ we get

$$\|\mathbf{a}_i\|_2^2 \geq = \beta T(\tfrac{1}{M} - \tfrac{1}{K+1}) \tag{28}$$

Next, consider the next link $(j, j+1)$ that is occupied by Eve and is upstream of link $(i, i+1)$. Now again, if the following holds

$$\|\mathbf{x}_{j+1}\|_2^2 - \|\mathbf{x}_j\|_2^2 > \tfrac{\beta}{K+1}T \tag{29}$$

then we are done, since link $(j, j+1)$ is adjacent to Eve. So, we again suppose (29) does *not* hold. Since Eve does not occupy any links between R_{j+1} and R_i, and only congestion-related loss could have occurred on the links between R_{j+1} and R_i. It follows that $\|\mathbf{x}_{j+1}\|_2^2 \geq \|\mathbf{x}_i\|_2^2 + \rho(i-j-1)$. Since (29) does *not* hold, we can apply identity (3) and the fact that $\|\mathbf{x}_{j+1}\|_2^2 \geq \|\mathbf{x}_i\|_2^2 + \rho(i-j-1) \geq \|\mathbf{a}_i\|_2^2 + \rho(i-j-1)$ and the bound on $\|\mathbf{a}_i\|_2^2$ in (28) to get

$$\|\mathbf{x}_j\|_2^2 > \beta T\left(\tfrac{1}{M} - \tfrac{2}{K+1} - \tfrac{\rho}{\beta}(i-j-1)\right)$$

We continue this argument for all $m \leq M-1$ links that are adjacent to Eve and upstream of link $(i, i+1)$. Finally, arriving at the last such link, which we call link $(e, e+1)$, we have

$$\begin{aligned}
\|\mathbf{x}_{e+1}\|_2^2 &> \beta T\left(\tfrac{1}{M} - \tfrac{m}{K+1} - \tfrac{\rho}{\beta}(i-e-1)\right) \\
&> \beta T\left(\tfrac{1}{M} - \tfrac{M-1}{K+1} - \tfrac{\rho}{\beta}K\right)
\end{aligned}$$

where the last inequality follows by putting $m \leq M-1$ and $i - e \leq K$. Now since by definition Eve does not occupy any links downstream of link $(e, e+1)$, we immediately have that $\|\mathbf{x}_e\|_2^2 = 0$. It follows that link $(e, e+1)$ has

$$\|\mathbf{x}_{e+1}\|_2^2 - \|\mathbf{x}_e\|_2^2 > \beta T(\tfrac{1}{M} - \tfrac{M-1}{K+1} - \tfrac{\rho}{\beta}K) > \tfrac{\beta}{K+1}$$

where the last inequality follows because we put $M \leq \sqrt{(K+1)(1-\frac{\rho}{\beta}K^2)}$. This concludes the proof of this lemma, since link $(e, e+1)$ is adjacent to Eve.

APPENDIX D
SKETCHING WITH PRFS

We now prove Theorem III.2. To do this, we first prove Theorem D.1, and then show how to derive Theorem III.2 from Theorem D.1. Recall from Section III-D that $\mathcal{S}$ is the set of $N \times |U|$ matrices where each column contains a single ± 1 entry in one row, and zeros in all other rows.

Theorem D.1. *For any vector $\mathbf{v} \in \mathbb{Z}^U$, choosing the $N \times U$ matrix S uniformly from $\mathcal{S}$ and setting $\mathbf{w} = S\mathbf{v}$, we have that for all $\varepsilon \in [0,1)$ and all $q, r > N$*

1) *If $\mathbf{v} \in \{-1,0,1\}^U$, and $\|v\|_2^2 \le q$, then for $\eta \in [0, \frac{1}{2}\sqrt{\varepsilon^2+10\varepsilon+9} - \frac{1}{2}(\varepsilon+3))$ and $y \doteq \frac{(1+\varepsilon)(1-\eta)}{(1+\eta)^2} - 1$:*

$$\Pr\left[\|\mathbf{w}\|_2^2 > (1+\varepsilon)q\right] \le 2Ne^{-\frac{\eta^2 q}{3N}} + e^{-\frac{N}{2}(y^2/2 - y^3/3)} \tag{30}$$

2) *If the number of* non-zero *entries in $\mathbf{v}$ is r, then for $\eta \in \left(0, \frac{1}{2-\varepsilon}(3-2\varepsilon-\sqrt{5\varepsilon^2-14\varepsilon+9})\right)$ and $y \doteq \frac{(1-\eta)^2}{1+\eta}(1-\frac{\varepsilon}{2}) - (1-\varepsilon)$ it follows that*

$$\Pr\left[\|\mathbf{w}\|_2^2 < (1-\varepsilon)r\right] \le 2Ne^{-\frac{\eta^2 r}{3N}} + e^{-N\frac{\varepsilon}{3(1+\eta)}y} \tag{31}$$

Proof of Theorem D.1. Our main observation is that, with high probability, the ± 1 entries of $\mathbf{v}$ are distributed evenly among the coordinates of $\mathbf{w}$. Conditioned on this happening, we can then apply the analysis of [1].

Definitions. We need the following definitions.

- We write v_x for the x^{th} element in $\mathbf{v}$.
- Define for $i \in [N]$ the set $Q_i = \{x \in U \mid h(x) = i\}$ where h is the pseudorandom hash function.
- Define D_i as the number of non-zero entries in $\mathbf{v}$ that hash to the i^{th} bin the sketch $\mathbf{w}$. That is $D_i = |\{v_x | v_x \neq 0, x \in Q_i\}|$.
- Define Y_x as an unbiased ± 1 random variable for each $x \in U$.

Our proof proceeds as follows. We first obtain a bound on D_i for each i. (Note: This bound on D_i gives rise to the awkward bound on T in Theorem III.2 of Section III-D.) When then use the bounds on D_i to prove the first item (30), and then use them to prove the second item (31).

Bounding D_i. Let E_i denote the event that $\exists i \in [N]$ such that $D_i > (1+\eta)q/N$ or $D_i < (1-\eta)q/N$. Then, for $\eta \in [0,1)$, we have that

$$\Pr[E_1] \le N\left(\Pr[D_i > (1+\eta)\tfrac{q}{N}] + \Pr[D_i < (1-\eta)\tfrac{q}{N}]\right) \le N\left(e^{-\frac{\eta^2}{3}\frac{q}{N}} + e^{-\frac{\eta^2}{2}\frac{q}{N}}\right) \tag{32}$$

which is a straightforward application of a union bound followed by the Chernoff bound.[6]

Bounding the first item. Now we condition on $\neg E_1$. Let $\gamma = \frac{1+\varepsilon}{(1+\eta)^2}$ and write:

$$\begin{aligned}\Pr[\|\mathbf{w}\|_2^2 > (1+\varepsilon)q \mid \neg E_1] &= \Pr[\sum_{i=1}^N D_i^2 \left(\tfrac{1}{D_i}\sum_{x\in Q_i} Y_x v_x\right)^2 > (1+\varepsilon)q \mid \neg E_1] \\ &= \Pr[\sum_{i=1}^N \left(\tfrac{1}{D_i}\sum_{x\in Q_i} Y_x v_x\right)^2 > \gamma\tfrac{N^2}{q} \mid \neg E_1]\end{aligned}$$

where first equality comes from expanding $\mathbf{w}$ as $\mathcal{S}\mathbf{v}$ and then multiplying by $\frac{D_i}{D_i}$, and the second equality follows from the fact that conditioning on $\neg E_i$ implies that $D_i \le (1+\eta)q/N$. Next, set $\mathbf{Y}_i$ to be the vector of all Y_x for each $v_x \in \{-1,1\}, x \in Q_i$. Set $\mathbf{u}_i$ the vector with entries $\frac{v_x}{\sqrt{D_i}}$ for each $v_x \in \{-1,1\}, x \in Q_i$. Notice that both $\mathbf{Y_i}$ and $\mathbf{u_i}$ have length D_i, and $\|\mathbf{u}_i\|_2^2 = 1$ so $\mathbf{u_i}$ is a unit vector. Now we write

$$\begin{aligned}&= \Pr[e^{t\sum_{i=1}^N \langle \frac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^2} > e^{t\gamma\frac{N^2}{q}} \mid \neg E_1] \\ &\le e^{-t\gamma\frac{N^2}{q}} \prod_{i=1}^N \mathbb{E}[e^{t\langle \frac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^2} \mid \neg E_1]\end{aligned}$$

[6]We use the following Chernoff bounds. Let X_i be i.i.d indicator variables with mean μ, and let

$$\Pr\left[\sum_{i=1}^n X_i \le (1-\gamma)N\mu\right] \le e^{-\gamma^2 N\mu/C_1}$$
$$\Pr\left[\sum_{i=1}^n X_i \ge (1+\gamma)N\mu\right] \le e^{-\gamma^2 N\mu/C_2}$$

If $0 < \gamma < 1$ then [5, Fact 4] gives $C_1 = 2$ and $C_2 = 3$. If $0 < \gamma < \frac{1}{2}$ then [2, Thm. 19] gives $C_1 = C_2 = 2\ln 2$.

where the inequality follows from the Markov bound. Now we are ready to apply the result of [1]. We restate equation (2) and Lemma 5.2 of [1] here, using our own terminology.

Lemma D.2 (From [1])**.** *For $t \in [0, D_i/2]$, unit vector $\mathbf{u}_i$ (*i.e., *$\|\mathbf{u}_i\|_2^2 = 1$) and $\mathbf{Y}_i$ chosen uniformly from $\{1,-1\}^{D_i}$ we have that*

$$\mathbb{E}[e^{t\langle \frac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^2}] \leq \frac{1}{\sqrt{1-2t/D_i}} \tag{33}$$

$$\mathbb{E}[\langle \tfrac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^2] = \frac{1}{D_i} \tag{34}$$

$$\mathbb{E}[\langle \tfrac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^4] = \frac{3}{D_i^2} \tag{35}$$

Now, using [1]'s result in (33) we write

$$\begin{aligned}
\Pr[\|\mathbf{w}\|_2^2 > (1+\varepsilon)q \mid \neg E_1] &\leq e^{-t\gamma\frac{N^2}{q}} \prod_{i=1}^{N} \mathbb{E}[e^{t\langle \frac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^2} \mid \neg E_1] \\
&\leq e^{-t\gamma\frac{N^2}{q}} \prod_{i=1}^{N} \frac{1}{\sqrt{1-2t/D_i}} \\
&\leq e^{-t\gamma\frac{N^2}{q}} (1-\tfrac{2t}{(1-\eta)q/N})^{-\frac{N}{2}} \doteq v(t)
\end{aligned} \tag{36}$$

where the last inequality (36) follows from conditioning on $\neg E_i$ which implies that $(1-\eta)q/N < D_i$ for all $i \in [N]$. Note that for the result of [1] in (33) to hold, we must have $0 \leq t < D_i/2 \leq \frac{(1+\eta)q}{2N}$ where the last inequality here follows from the fact that $\neg E_i$ implies that $D_i < (1+\eta)q/N$.

Optimizing and bounding t. Next, we optimize $v(t)$ in (36), by finding t such that $\frac{dv(t)}{dt} = 0$.

$$\begin{aligned}
\tfrac{dv(t)}{dt} = -\tfrac{\gamma N^2}{q} v(t) + (-\tfrac{N}{2})(-\tfrac{2}{(1-\eta)q/N})(1-\tfrac{2t}{(1-\eta)q/N})^{-1} v(t) &= 0 \\
\tfrac{\gamma N^2}{q}(1-\tfrac{2t}{(1-\eta)q/N}) &= \tfrac{N^2}{(1-\eta)q} \\
t &= \tfrac{q}{2N}\left((1-\eta) - \tfrac{(1+\eta)^2}{1+\varepsilon}\right)
\end{aligned} \tag{37}$$

where the last equality uses the fact that $\gamma \doteq \frac{1+\varepsilon}{(1+\eta)^2}$. Now recall that for [1]'s result in (33) to hold, we need to ensure that $0 \leq t < \frac{(1+\eta)q}{2N}$. Using (37), we write

$$\begin{aligned}
0 &\leq t \\
0 &\leq \tfrac{q}{2N}\left((1-\eta) - \tfrac{(1+\eta)^2}{1+\varepsilon}\right) \\
\tfrac{(\eta^2+3\eta)}{1-\eta} &\leq \varepsilon
\end{aligned} \tag{38}$$

and we also need

$$\begin{aligned}
t &< \tfrac{(1+\eta)q}{2N} \\
\tfrac{q}{2N}\left((1-\eta) - \tfrac{(1+\eta)^2}{1+\varepsilon}\right) &< \tfrac{(1+\eta)q}{2N} \\
-\left(1 + \tfrac{(1+\eta)^2}{2\eta}\right) &< \varepsilon
\end{aligned} \tag{39}$$

Now, (39) holds for any $\eta \in [0,1)$. But, we will need to ensure that our choice of $\eta \in [0,1)$ satisfies (38).

Returning now to (36), plug (37) into (36) to get

$$\Pr[\|\mathbf{w}\|_2^2 > (1+\varepsilon)q \mid \neg E_1] \leq \left(e^{-y}(1+y)\right)^{\frac{N}{2}} \tag{40}$$

where we define

$$y \doteq \frac{(1+\varepsilon)(1-\eta)}{(1+\eta)^2} - 1 \tag{41}$$

and solving inequality (38), we find that (40) holds as long as $\eta \in [0,1)$ satisfies

$$0 < \eta < \tfrac{1}{2}\left(\sqrt{\varepsilon^2+10\varepsilon+9} - (\varepsilon+3)\right) \tag{42}$$

Notice from (41) that the bound in (42) this implies that (40) holds for the region $y \in [0, \varepsilon)$. Now, [1] observes that $e^{-y}(1+y) \leq e^{(-y^2/2+y^3/3)}$ for any $y \in (0,1)$. Since for us $y \in (0,\varepsilon)$, and $\varepsilon < 1$ we finally have

$$\Pr[\|\mathbf{w}\|_2^2 > (1+\varepsilon)q \mid \neg E_1] \leq e^{-\frac{N}{2}(y^2/2-y^3/3)} \tag{43}$$

which decays exponentially in N. □

Bounding the second item. Let r be the number of non-zero entries in $\mathbf{v}$. We will bound $\Pr[\|\mathbf{w}\|_2^2 < (1-\varepsilon)r]$. Define E_1 as before, only this time use r instead of q. Again we condition on $\neg E_1$.

$$\begin{aligned}\Pr[\|\mathbf{w}\|_2^2 < (1-\varepsilon)r \mid \neg E_1] &= \Pr[\sum_{i=1}^{N} D_i^2 \left(\tfrac{1}{D_i}\sum_{x\in Q_i} Y_x v_x\right)^2 < (1-\varepsilon)r \mid \neg E_1] \\ &= \Pr[\sum_{i=1}^{N} \left(\tfrac{1}{D_i}\sum_{x\in Q_i} Y_x v_x\right)^2 < \tfrac{(1-\varepsilon)}{(1-\eta)^2}\tfrac{N^2}{r} \mid \neg E_1]\end{aligned}$$

where first equality comes from the expanding $\|\mathbf{w}\|_2^2$and then multiplying by $\frac{D_i}{D_i}$, and the second equality follows from the fact that conditioning on $\neg E_i$ implies that $(1-\eta)r/N < D_i$. Next, we let $c_i^2 = \sum_{x\in Q_i} \frac{v_x^2}{D_i}$. Now observe that $c_i^2 = \frac{1}{D_i}\sum_{x\in Q_i} v_x^2 \geq \frac{1}{D_i} D_i = 1$ since the entries of v are integers (and D_i is the number of non-zero entries in v that are in Q_i). We now multiply by $\frac{c_i}{c_i}$:

$$\begin{aligned}&= \Pr[\sum_{i=1}^{N} c_i^2 \left(\sum_{x\in Q_i} Y_x \tfrac{v_x}{D_i c_i}\right)^2 < \tfrac{(1-\varepsilon)}{(1-\eta)^2}\tfrac{N^2}{r} \mid \neg E_1] \\ &\leq \Pr[\sum_{i=1}^{N} \left(\sum_{x\in Q_i} Y_x \tfrac{v_x}{D_i c_i}\right)^2 < \tfrac{(1-\varepsilon)}{(1-\eta)^2}\tfrac{N^2}{r} \mid \neg E_1]\end{aligned}$$

where the inequality follows from the fact that $c_i^2 \geq 1$. We now set $\mathbf{Y}_i$ to be the vector of all Y_x for each $v_x \neq 0, x \in Q_i$. Set $\mathbf{u}_i$ the vector with entries $\frac{v_x}{\sqrt{D_i}c_i}$ for each $v_x \neq 0, x \in Q_i$. Notice that both $\mathbf{Y}_i$ and $\mathbf{u}_i$ have length D_i, and that $\mathbf{u}_i$ is a unit vector, since $\|\mathbf{u}_i\|_2^2 = \frac{1}{D_i c_i^2}\sum_{x\in Q_i} v_x = \frac{c_i^2}{c_i^2} = 1$. We write

$$\begin{aligned}&= \Pr[\sum_{i=1}^{N} \langle \tfrac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^2 < \tfrac{(1-\varepsilon)}{(1-\eta)^2}\tfrac{N^2}{r} \mid \neg E_1] \\ &\leq e^{t\frac{(1-\varepsilon)}{(1-\eta)^2}\frac{N^2}{r}} \prod_{i=1}^{N} \mathbb{E}[e^{-t\langle \frac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^2} \mid \neg E_1]\end{aligned}$$

where the first inequality follows from the Markov bound, and we require that $t > 0$. We now follow that analysis in Achioptas, and expand out the quantity inside the expectation to obtain:

$$\leq e^{t\frac{(1-\varepsilon)}{(1-\eta)^2}\frac{N^2}{r}} \prod_{i=1}^{N} \mathbb{E}[1 - t\langle \tfrac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^2 + \tfrac{t^2}{2}\langle \tfrac{\mathbf{Y}_i}{\sqrt{D_i}}, \mathbf{u}_i\rangle^4 \mid \neg E_1]$$

Now we can apply Achioptas's results from (34) and (35) to obtain:

$$\leq e^{t\frac{(1-\varepsilon)}{(1-\eta)^2}\frac{N^2}{r}} \prod_{i=1}^{N} \left(1 - \tfrac{t}{D_i} + \tfrac{t^2}{2}\tfrac{3}{D_i^2}\right)$$

and conditioning on $\neg E_1$ gives us:

$$\leq e^{t\frac{(1-\varepsilon)}{(1-\eta)^2}\frac{N^2}{r}} \left(1 - \tfrac{1}{1+\eta}\tfrac{tN}{r} + \tfrac{3}{2(1-\eta)^2}(\tfrac{tN}{r})^2\right)^N$$

For convience, we'll now let $\tau = \frac{tN}{r}$, and rewrite this as

$$= \left(e^{\frac{(1-\varepsilon)}{(1-\eta)^2}\tau}\left(1 - \tfrac{1}{1+\eta}\tau + \tfrac{3}{2(1-\eta)^2}\tau^2\right)\right)^N \doteq \nu(\tau)^N \tag{44}$$

Bounding equation (44). We now need to find a choice of $\tau > 0$ that causes (44) to decay with N. It will suffice to find τ that causes $\nu(\tau)$ to decay exponentially, *i.e.,* we want $\nu(\tau) \sim e^{-\chi}$ for some $\chi > 0$. To do this, we start by rewriting $\nu(\tau)$ in the following way:

$$\nu(\tau) = e^{\frac{(1-\varepsilon)}{(1-\eta)^2}\tau}\left(1 - \tfrac{1}{1+\eta}\tau \cdot \left(1 - \tfrac{3}{2}\tfrac{(1+\eta)}{(1-\eta)^2} \cdot \tau\right)\right)$$

Notice that $\nu(\tau)$ is the product of a polynomial and exponential with postive argument (that grows). Notice that the only way we can hope to make $\nu(\tau)$ decay, is if we require the polynomial to decay. To do this, we need to ensure that the expression $(1 - \frac{3}{2}\frac{(1+\eta)}{(1-\eta)^2} \cdot \tau)$ is positive. Thus, we shall choose $\tau = \frac{\varepsilon}{2}(\frac{3}{2}\frac{(1+\eta)}{(1-\eta)^2})^{-1}$. Subsituting in the value for τ gives us:

$$= e^{\frac{1-\varepsilon}{1+\eta}\frac{\varepsilon}{3}}\left(1 - (\tfrac{1-\eta}{1+\eta})^2\tfrac{\varepsilon}{3} \cdot (1 - \tfrac{\varepsilon}{2})\right)$$

The series expansion of an exponential tell us that for any non-negative x we have the identity $1 - x \leq e^{-x}$. Since the quantity $(\frac{1-\eta}{1+\eta})^2\frac{\varepsilon}{3} \cdot (1 - \frac{\varepsilon}{2})$ is non-negative for every $\varepsilon \in (0, 1)$, we can apply this identity here:

$$\begin{aligned} &\leq \exp\left(\tfrac{1-\varepsilon}{1+\eta}\tfrac{\varepsilon}{3} - (\tfrac{1-\eta}{1+\eta})^2\tfrac{\varepsilon}{3} \cdot (1 - \tfrac{\varepsilon}{2})\right) \\ &= e^{-\frac{\varepsilon}{3(1+\eta)}}\exp\left(\tfrac{(1-\eta)^2}{1+\eta}(1 - \tfrac{\varepsilon}{2}) - (1 - \varepsilon)\right) \end{aligned} \tag{45}$$

It follows from (45) that proving that $\nu(\tau)$ decays exponentially amounts to ensuring that

$$y(\eta, \varepsilon) \doteq \frac{(1-\eta)^2}{1+\eta}(1 - \tfrac{\varepsilon}{2}) - (1 - \varepsilon) \geq 0 \tag{46}$$

and, recalling that $\eta, \varepsilon \in (0, 1)$ some algebraic manipulation finds that (46) holds as long as $\eta \in (0, c(\varepsilon))$, where

$$c(\varepsilon) = \tfrac{1}{2-\varepsilon}(3 - 2\varepsilon - \sqrt{5\varepsilon^2 - 14\varepsilon + 9}) \tag{47}$$

This bound on η, despite being ugly, makes sense. Notice that when $\varepsilon = 0$, we have that $\eta = 0$, and when $\varepsilon = 1$, we have $c(\varepsilon) = 1$ so that $\eta \in (0, 1)$. Also, we observe that y monotonically decreases in η, ranging from $y(0, \varepsilon) = \varepsilon$ to $y((c(\varepsilon), \varepsilon) = 0$.[7] We also observe that y monotonically increase in ε, ranging from $y(\eta, 0) = y(0, 0) = 0$ (since $\eta = 0$ when $\varepsilon = 0$), and $y(\eta, 1) = \frac{1}{2}\frac{(1-\eta)^2}{1+\eta}$ (and $\eta \in (0, 1)$ when $\varepsilon = 1$). [8]

Putting everything together, we finally have that as long as $\eta \in (0, c(\varepsilon))$ where $c(\varepsilon)$ is given in (47), then y as given in (46) is such that $y > 0$. Re-writing (44) using (45) and (46) as

$$\Pr[\|\mathbf{w}\|_2^2 < (1 - \varepsilon)r \mid \neg E_1] \leq e^{-N\frac{\varepsilon}{3(1+\eta)}y} \tag{48}$$

we can see that the error decays exponentially in N, as required. □

A simpler statement of the theorem. We now prove Theorem III.2 from Theorem D.1.

Proof of Theorem III.2. We show how to obtain the Theorem III.2 from Theorem D.1. To ensure that the error probability is at most δ in (30) it suffices to set

$$2Ne^{-\frac{\eta^2 q}{3N}} \leq \tfrac{\delta}{2} \tag{49}$$

$$e^{-\frac{N}{2}(y_1^2/2 - y_1^3/3)} \leq \tfrac{\delta}{2} \tag{50}$$

And to ensure that the error probability is at most δ in (31) we need to set

$$2Ne^{-\frac{\eta^2 r}{3N}} \leq \tfrac{\delta}{2} \tag{51}$$

$$e^{-N\frac{\varepsilon}{3(1+\eta)}y} \leq \tfrac{\delta}{2} \tag{52}$$

Bounding N. Referring to (50), we need to choose $N > N_{\min,1}$ where:

$$N_{\min,1} = \frac{4}{y_1^2(1 - y_1/6)}\ln\tfrac{2}{\delta} \tag{53}$$

[7]One can see that when $\eta = 0$, then $y(0, \varepsilon) = \varepsilon$, and a simple check in MATHEMATICA shows that when $\eta = c(\varepsilon)$ as in (47), then $y(c(\varepsilon), \varepsilon) = 0$. By inspection, it follows that y decreases in η.

[8]First consider the case where $\varepsilon = 0$. Now when $\varepsilon = 0$, $c(\varepsilon) = 0$, and the requirement that $\eta \in (0, c(\varepsilon))$ implies that $\eta = 0$. It follows that $y = 0$. Next consider the case where $\varepsilon = 1$, which means that for $\eta \in (0, 1)$, we have that $y(\eta, 0) = \frac{1}{2}\frac{(1-\eta)^2}{1+\eta}$. Now, since the derivative $\frac{dy}{d\varepsilon} = \frac{1+\eta(4-\eta)}{1+\eta} > 0$ for any $\eta \in (0, 1)$, we know that y grow monotonically in ε.

Where recall that $y_1 \doteq \frac{(1+\varepsilon)(1-\eta)}{(1+\eta)^2} - 1$. One can verify that $y_1 \in (0, \varepsilon)$ for any $\eta, \varepsilon \in (0,1)$. To simplify (53), we will now require that $y_1 \geq \varepsilon/2$, which means we can write:

$$\begin{aligned} &\leq \frac{4}{y_1^2(1-\varepsilon/6)} \ln \tfrac{2}{\delta} \\ &\leq \frac{4}{(\varepsilon/2)^2(1-\varepsilon/6)} \ln \tfrac{2}{\delta} \\ &\leq \frac{19.2}{\varepsilon^2} \ln \tfrac{2}{\delta} \end{aligned}$$

where the first inequality follows because $y \leq \varepsilon$, the second follows from $y \geq \varepsilon/2$, and the third follows from $\varepsilon \leq 1$. Now, instead of using the "ugly" expression for $N > N_{\min,1}$ in (53) to bound N, we have "nicer" bound on N that shows the dependence of N on ε, δ as:

$$N \geq \frac{19.2}{\varepsilon^2} \ln \tfrac{2}{\delta} \tag{54}$$

Next, refer to (52), we need to choose $N > N_{\min,2}$ where:

$$N_{\min,2} = \tfrac{3(1+\eta)}{\varepsilon y_2} \ln \tfrac{2}{\delta} \tag{55}$$

Where recall that $y_2 = \frac{(1-\eta)^2}{1+\eta}(1-\frac{\varepsilon}{2}) - (1-\varepsilon)$. One can see that $y_2 \in (0, \frac{\varepsilon}{2})$ for any $\eta \in (0,1)$. To simplify (53), we will now require that $y_2 \geq \varepsilon/4$ which means we can write:

$$\begin{aligned} &\leq \tfrac{12(1+\eta)}{\varepsilon^2} \ln \tfrac{2}{\delta} \\ &\leq \tfrac{24}{\varepsilon^2} \ln \tfrac{2}{\delta} \end{aligned}$$

where the first inequality follows from our choice of $y_2 \geq \varepsilon/4$ and the second from $\eta \leq 1$. Now we again have "nicer" bound on N (showing it's dependence of N on ε, δ) as:

$$N \geq \tfrac{24}{\varepsilon^2} \ln \tfrac{2}{\delta} \tag{56}$$

Comparing equations (54) and (56) we find that it suffices to choose N satisfying (56).

Bounding η. These nice bounds on N does not come free. To obtain (54), we need to ensure that $y_1 > \varepsilon/2$. We write

$$\begin{aligned} \tfrac{\varepsilon}{2} \leq y_1 &\doteq \tfrac{(1+\varepsilon)(1-\eta)}{(1+\eta)^2} - 1 \\ \tfrac{1+\frac{\varepsilon}{2}}{1+\varepsilon} &\leq \tfrac{1-\eta}{(1+\eta)^2} \end{aligned} \tag{57}$$

Now since $\frac{1+\frac{\varepsilon}{2}}{1+\varepsilon} \leq \left(\frac{1-\eta}{1+\eta}\right)^2 \leq \frac{1-\eta}{(1+\eta)^2}$ it follows that (57) holds if

$$\tfrac{1+\frac{\varepsilon}{2}}{1+\varepsilon} \leq \left(\tfrac{1-\eta}{1+\eta}\right)^2 \tag{58}$$

Next, to obtain (56) we need ensure that $y_2 > \varepsilon/4$, so we write

$$\tfrac{\varepsilon}{4} \leq y_2 \doteq \tfrac{(1-\eta)^2}{1+\eta}(1-\tfrac{\varepsilon}{2}) - (1-\varepsilon) \tag{59}$$

and a similar argument show that (59) holds as long as

$$\tfrac{1-\frac{3\varepsilon}{4}}{1-\frac{\varepsilon}{2}} \leq \left(\tfrac{1-\eta}{1+\eta}\right)^2 \tag{60}$$

Bounding q, r. Referring to (49) and (51), we observe that is suffices to choose

$$q, r \geq \tfrac{3N}{\eta^2} \ln \tfrac{4N}{\delta} \tag{61}$$

Notice that this bound relies on both N, and η. We bounded N in (56). To minimize q, r, we want to chose η as large as possible, subject to the constraints in (58) and (60). Thus, it suffices to chose η such that

$$\left(\tfrac{1-\eta}{1+\eta}\right)^2 = \max\left(\tfrac{1+\frac{\varepsilon}{2}}{1+\varepsilon},\ \tfrac{1-\frac{3\varepsilon}{4}}{1-\frac{\varepsilon}{2}}\right) \tag{62}$$

and this completes our proof of Theorem III.2. □