Dandelion-Lite: Protocol and Analysis

November 5, 2018

Abstract

This note analyzes a simplified version of Dandelion called Dandelion-Lite, in which the stem phase always lasts exactly one hop. We show that under the assumptions that the adversary already knows the underlying graph topology, Dandelion-Lite gives similar privacy guarantees to Dandelion, while being substantially simpler.

1 Protocol

We consider a simplified version of Dandelion. We call it Dandelion-Lite, or Lite for short, and summarize it below:

- Each node v chooses one outbound edge to a peer w for Dandelion transactions; we call w the Dandelion relay
- Upon generating a transaction, the source v forwards its transaction *only* to the Dandelion relay w, *without* using a separate message type. It also starts a random timer.
- The recipient of any transaction relays the message as a normal transaction
- Sequential transactions generated by the same node are sent to the same Dandelion relay
- If v's timer expires before it receives an `INV` for the transaction from a node other than the Dandelion relay, it starts the fluff phase.

2 Analysis

We analyze two metrics: expected precision and recall. The model is the same as in our prior papers—we allow each node to send multiple transactions, and assume the adversary can link them together. The following proposition shows that when the adversary knows the graph, Lite has an expected optimal precision and recall that are both $p + o(1)$. This is comparable to Dandelion; when the adversary knows the graph, it has an expected precision that scales as p [2, Proposition 1]. Although the result in [2] is an upper bound (i.e., precision scales as $O(p)$), we observe empirically that the precision is indeed larger than p in simulation. Hence, Lite does not appear to significantly weaken privacy gains when the adversary learns the graph compared to Dandelion, but it comes with a significant reduction in implementation complexity.

Proposition 1.

$$\mathbb{E}[\mathbf{R}_{\texttt{OPT}}] \leq p + \frac{d(1-p)}{n} \tag{1}$$

$$\mathbb{E}[\mathbf{D}_{\texttt{OPT}}] \leq p + \frac{d(1-p)}{n} \tag{2}$$

Proof. We first prove the result for recall, then precision.

Recall. From Theorem 4 of [1], we know that a recall-optimal estimator for any spreading mechanism is the first-spy estimator. Hence we upper bound the expected recall of that estimator. Let R_x denote the event where transaction x initially gets sent to a spy node, and $\bar{R}_x$ the complement of that event. We assume the graph is formed by each honest node choosing d outbound edges uniformly at random from the other nodes. This gives

$$\begin{aligned}\mathbb{E}[\mathbf{R}_{\texttt{OPT}}] &\leq \mathbb{P}(R_x) \cdot \mathbb{E}[\mathbf{R}_{\texttt{OPT}}|R_x] + \mathbb{P}(\bar{R}_x) \cdot \mathbb{E}[\mathbf{R}_{\texttt{OPT}}|\bar{R}_x] \\ &= p \cdot 1 + (1-p) \cdot \mathbb{E}[\mathbf{R}_{\texttt{OPT}}|\bar{R}_x].\end{aligned}$$

To upper bound $\mathbb{E}[\mathbf{R}_{\texttt{OPT}}|\bar{R}_x]$, we assume that the adversary can always perfectly identify w, the fluff source (i.e., the node who starts diffusing the true source's transaction). Given this information, an ML estimator is to choose uniformly among the inbound neighbors of w, which we denote $\partial_{in}(w)$. The expected accuracy of this estimator is $\mathbb{E}[\frac{1}{\partial_{in}(w)}]$, where the expectation is taken over randomness in the graph topology, the source node v, and the spreading decision. Let $q = 1 - \left(\frac{n-1}{n}\right)^d$, where recall d is the out-degree of each honest node.

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\partial_{in}(w)}\right] &\leq \sum_{k=1}^{n} \frac{1}{k} \cdot q^k (1-q)^{n-k} \\ &\leq \sum_{k=1}^{\infty} \frac{1}{k} \cdot q^k \\ &= -\log(1-q) = -d\log(1-\frac{1}{n}) \\ &\leq \frac{d}{n}.\end{aligned}$$

This gives

$$\mathbb{E}[\mathbf{R}_{\texttt{OPT}}] \leq p \cdot 1 + (1-p)\frac{d}{n} = p + o(1),$$

where the asymptotics are taken as $n \to \infty$.

Precision. Precision follows from the following theorem, reproduced from [1] for convenience.

Theorem 2 (Theorem 1 from [1])**.** *Any mapping policy $\texttt{M} \in \mathcal{M}_{\tau,\sigma}$ on a network with topology $\tau \in \mathcal{T}$ and spreading strategy $\sigma \in \Sigma$ has a precision and recall that are bounded as*

$$\mathbf{D}_{\texttt{M}} \overset{(a)}{\leq} \mathbf{R}_{\texttt{M}} \overset{(b)}{\leq} \sqrt{\mathbf{D}_{\texttt{M}}}. \tag{3}$$

Hence the optimal precision is upper bounded by the optimal recall, which, combined with (1), gives the result.

□

References

[1] Shaileshh Bojja Venkatakrishnan, Giulia Fanti, and Pramod Viswanath. Dandelion: Redesigning the bitcoin network for anonymity. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 1(1):22, 2017.

[2] Giulia Fanti, Shaileshh Bojja Venkatakrishnan, Surya Bakshi, Bradley Denby, Shruti Bhargava, Andrew Miller, and Pramod Viswanath. Dandelion++: Lightweight cryptocurrency networking with formal anonymity guarantees. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2(2):29, 2018.