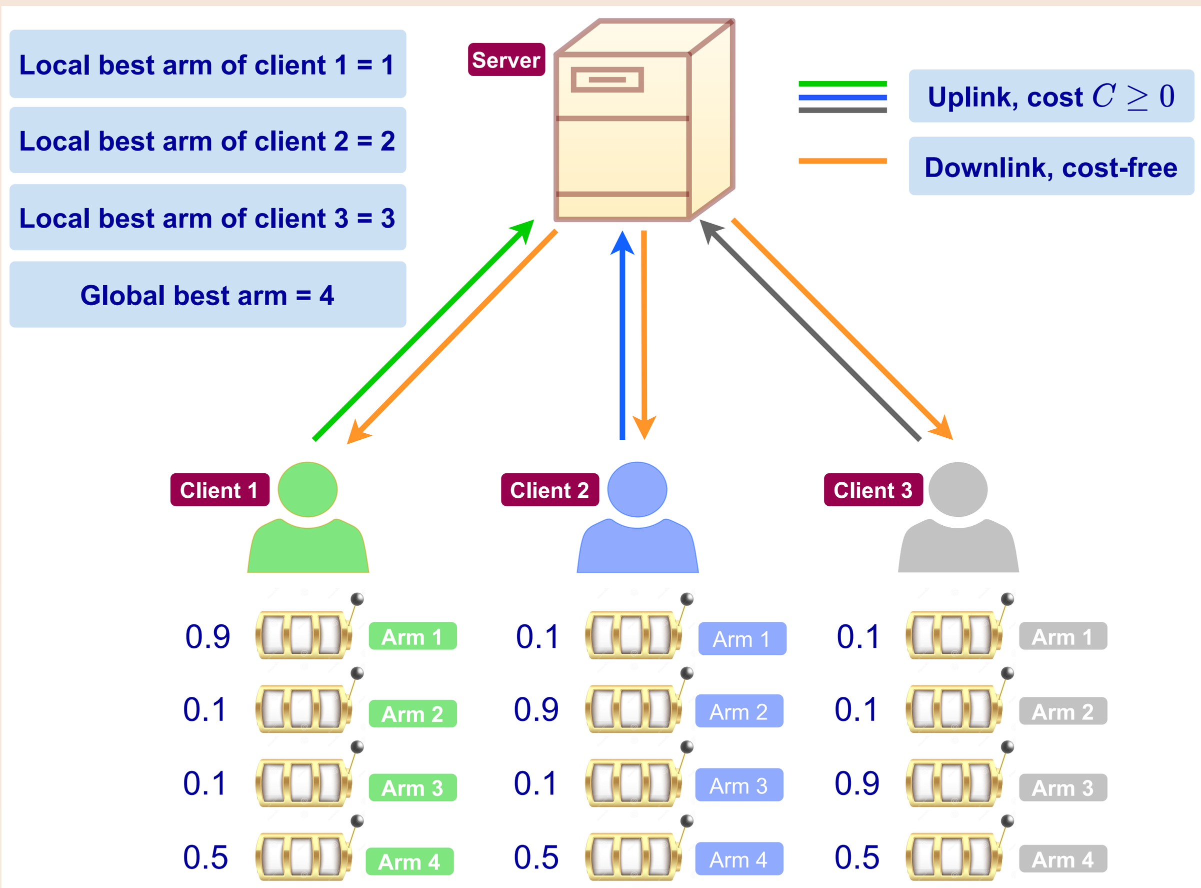
Almost Cost–Free Communication in Federated Best Arm Identification

Kota Srinivas Reddy **P. N. Karthik** Vincent Y. F. Tan

National University of Singapore

Contact: ksvr1532@gmail.com, karthik@nus.edu.sg, vtan@nus.edu.sg

Problem Setup



- M clients, one central server.
- K-armed bandit at each client.
- $\mu_{k,m}$: mean reward of arm k at client m.
- *Local best arm* of client $m = \arg\max_k \mu_{k,m}$.
- *Global best arm* $= \arg\max_k \frac{1}{M}\sum_k \mu_{k,m}$.
- Uplink cost $= C \geq 0$, Downlink = cost-free.
- Goal: Find the local best arms and global best arm by minimising
  total cost = no. of arm pulls + comm. cost,
  subject to upper bound on error probability.

Key Takeaways

- Total cost of FedElim when $C > 0$ is at most 3 times that of FedElim0, independent of C. That is, *communication is almost cost-free*.
- FedElim diligently balances no. of arm pulls and comm. cost, *without knowing C*.

Full Text

Scan to access the full text.
arXiv ID: 2208.09215.

Results at a Glance

Notation: $\boldsymbol{\mu} = \{\mu_{k,m} : k \in [K], m \in [M]\}$, k_m^*–local best arm of client m, $m \in [M]$, k^*–global best arm
$\hat{\mu}_{k,m}(n)$ – empirical estimate of the mean of arm k of client m in round n

$$\Delta_{k,m} := \begin{cases} \mu_{k_m^*,m} - \mu_{k,m}, & k \neq k_m^*, \\ \min_{k \neq k_m^*} \mu_{k_m^*,m} - \mu_{k,m}, & k = k_m^*, \end{cases} \qquad \Delta_k := \begin{cases} \frac{1}{M}\sum_{m=1}^M (\mu_{k^*,m} - \mu_{k,m}), & k \neq k^*, \\ \min_{k \neq k^*} \frac{1}{M}\sum_{m=1}^M (\mu_{k^*,m} - \mu_{k,m}), & k = k^*. \end{cases}$$

$$\mathcal{E} = \bigcap_{n=1}^{\infty}\bigcap_{k=1}^{K}\bigcap_{m=1}^{M} \left\{ \begin{matrix} |\hat{\mu}_k(n) - \mu_k| \leq \alpha_{\rm g}(n), \\ |\hat{\mu}_{k,m}(n) - \mu_{k,m}| \leq \alpha_{\rm l}(n) \end{matrix} \right\}, \quad T_{k,m} = O\left(\frac{\ln\left(\frac{64\sqrt{\frac{8KM}{\delta}}}{\Delta_{k,m}^2}\right)}{\Delta_{k,m}^2} \right), \quad T_k = O\left(\frac{\ln\left(\frac{64\sqrt{\frac{8K}{\delta}}}{M\Delta_k^2}\right)}{M\Delta_k^2} \right).$$

Theorem 1 *Fix $\boldsymbol{\mu}$, $\delta \in (0,1)$, and let $C = 0$.*

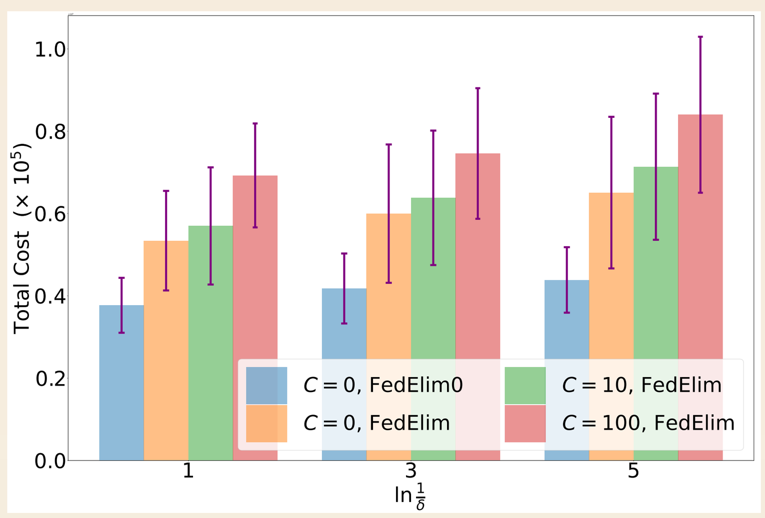
1. $P(\mathcal{E}) \geq 1 - \delta$.
2. *On the event $\mathcal{E}$, the total no. of arm pulls under FedElim0 is upper bounded by*
$$T = \sum_{k=1}^K \sum_{m=1}^M \max\{T_{k,m}, T_k\}.$$
3. *For any δ-PAC algorithm, the expected no. of arm pulls is lower bounded by*
$$\sum_{k=1}^K \sum_{m=1}^M \max\left\{ \frac{\ln(\frac{1}{2.4\delta})}{\Delta_{k,m}^2}, \frac{\ln(\frac{1}{2.4\delta})}{M^2\Delta_k^2} \right\}.$$

Theorem 2 *Fix $\boldsymbol{\mu}$, uplink cost $C \geq 0$, and $\delta \in (0,1)$ such that $C \ln T_k \leq T_k$ for all $k \in [K]$. Let $T^C_{\rm FedElim}$, $C^{\rm comm}_{\rm FedElim}$, and $C^{\rm total}_{\rm FedElim}$ denote respectively the total no. of arm pulls, the communication cost, and the total cost of FedElim towards identifying the local best arms and the global best arm. On the event $\mathcal{E}$ defined above, the following hold:*

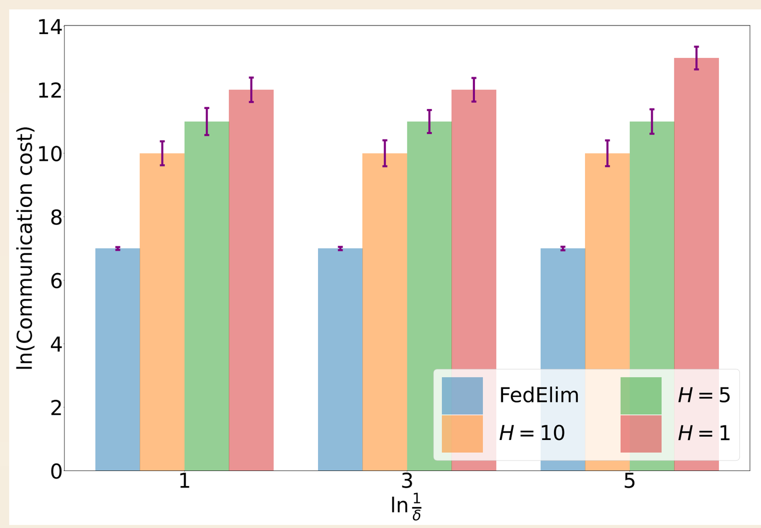
$$T^C_{\rm FedElim} \leq \sum_{k=1}^K \sum_{m=1}^M \max\{T_{k,m},\ 2T_k\} \leq 2T,$$

$$C^{comm}_{\rm FedElim} \leq C \cdot M \cdot \sum_{k=1}^K \left\lceil \frac{\ln T_k}{\ln 2} \right\rceil,$$

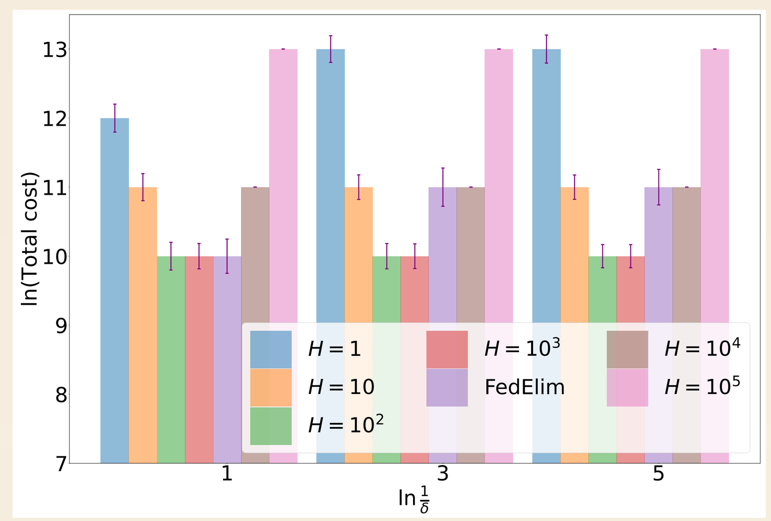
$$C^{total}_{\rm FedElim} = T^C_{\rm FedElim} + C^{\rm comm}_{\rm FedElim} \leq 3T.$$



Comparison of no. of arm pulls of FedElim0 with the total cost of FedElim for $C \in \{0, 10, 100\}$. FedElim0 has a lower cost compared to FedElim when $C = 0$.



Comparison of the communication cost incurred under FedElim and under variants of FedElim communicating periodically with period $H \in \{1, 5, 10\}$, with $C = 10$.



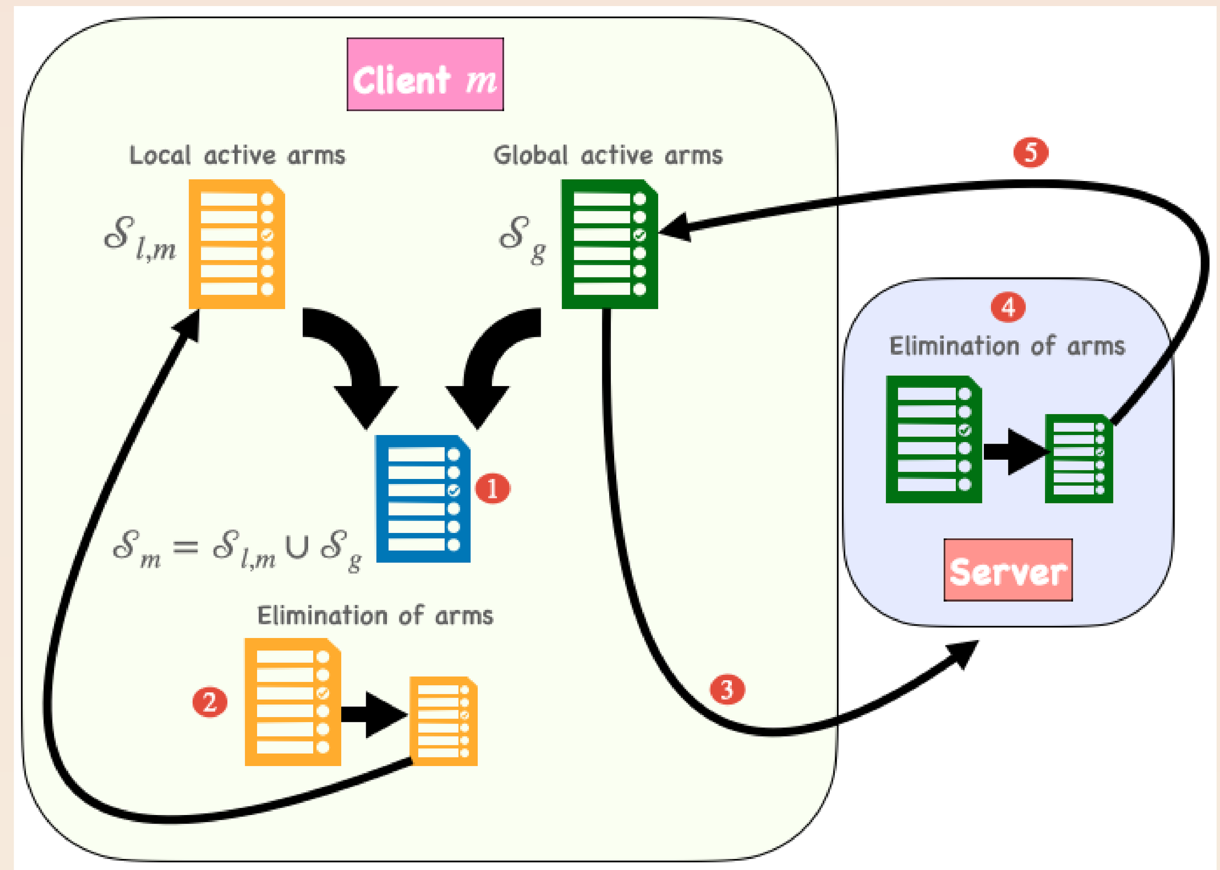
Comparison of the total cost under FedElim and under periodic communication with period $H = 10^p$ for $p \in \{0, 1, 2, 3, 4, 5\}$ and $C = 10$. FedElim enjoys a "sweet spot".

Simulation instance ($M = 3$, $K = 4$):

$$\boldsymbol{\mu} = \begin{bmatrix} 0.9 & 0.1 & 0.1 \\ 0.1 & 0.9 & 0.1 \\ 0.1 & 0.1 & 0.9 \\ 0.5 & 0.5 & 0.5 \end{bmatrix}.$$

Scheme	No. of Arm Pulls	Comm. Cost	Total Cost
Per(H)	$O(T)$	$O(CT/H)$	$O((1+C/H)T)$
FedElim	$2 \cdot T$	$O(C \ln T)$	$3 \cdot T$
Super-Exponential	$O(T^2)$	$O(C \ln\ln T)$	$O(T^2)$

FedElim0 ($C = 0$), FedElim ($C > 0$)

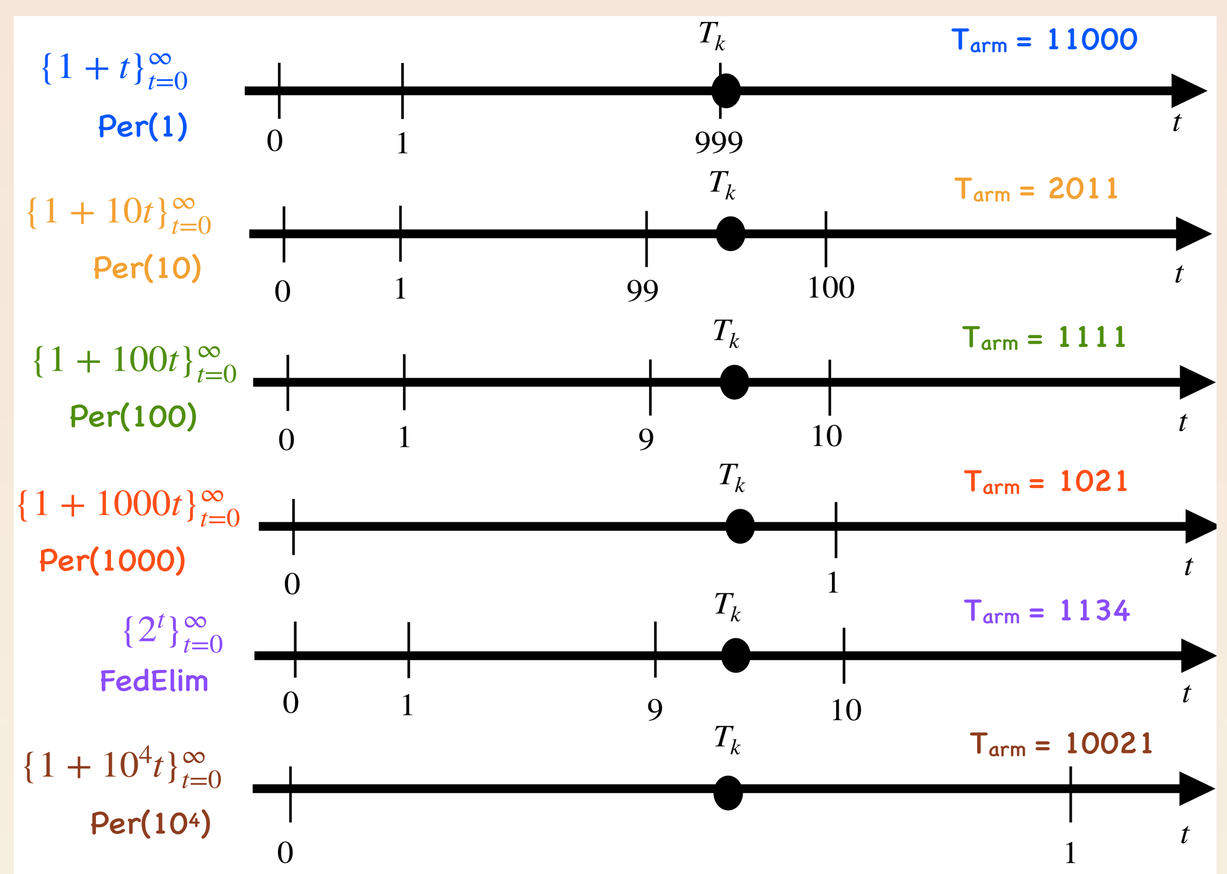


0. Initialise $\mathcal{S}_{l,m} = [K]$, $\mathcal{S}_m = [K]$ for all m
1. Pull $k \in \mathcal{S}_m$ and update $\hat{\mu}_{k,m}(n)$
2. If $\hat{\mu}_{k,m}(n) \leq \max_a \hat{\mu}_{a,m}(n) - 2\alpha_{\rm l}(n)$, eliminate k from $\mathcal{S}_{l,m}$
3. Client m sends $\hat{\mu}_{k,m}(n)$, $k \in \mathcal{S}_g$, if $n \in \{2^t\}_{t=0}^{\infty}$ (resp. $\forall n$) and $C > 0$ (resp. $C = 0$)
4. If $\hat{\mu}_{k,m}(n) \leq \max_a \hat{\mu}_{a,m}(n) - 2\alpha_{\rm g}(n)$, server eliminates k from $\mathcal{S}_g$
5. Server sends updated $\mathcal{S}_g$ to all clients

Elimination continues until $|\mathcal{S}_{l,m}| = 1\, \forall m$, $|\mathcal{S}_g| = 1$

$$\alpha_{\rm l}(n) = \sqrt{\frac{2\ln(8KMn^2/\delta)}{n}}, \quad \alpha_{\rm g}(n) = \sqrt{\frac{2\ln(8Kn^2/\delta)}{Mn}}.$$

FedElim vs Per(H)



$C = 10$ $T_k = 1000$ $T_{\rm arm}$ = no. of pulls + $10(t+1)$