The impact of crash tolerance on consensus time for blockchains

Y.C. Tay

(keynote for *IWSF&SHIFT 2025*, Sao Paolo, Brazil)

The impact of crash tolerance on consensus time for blockchains

Y.C. Tay

(keynote for *IWSF&SHIFT 2025*, São Paulo, Brazil)

objective: analytical model that expresses performance in terms of protocol and topology parameters

Medium   Search   Write

HotStuff: the Consensus Protocol Behind Facebook's LibraBFT

The Ontology Team · Follow
Published in OntologyNetwork · 7 min read · Jun 26, 2019

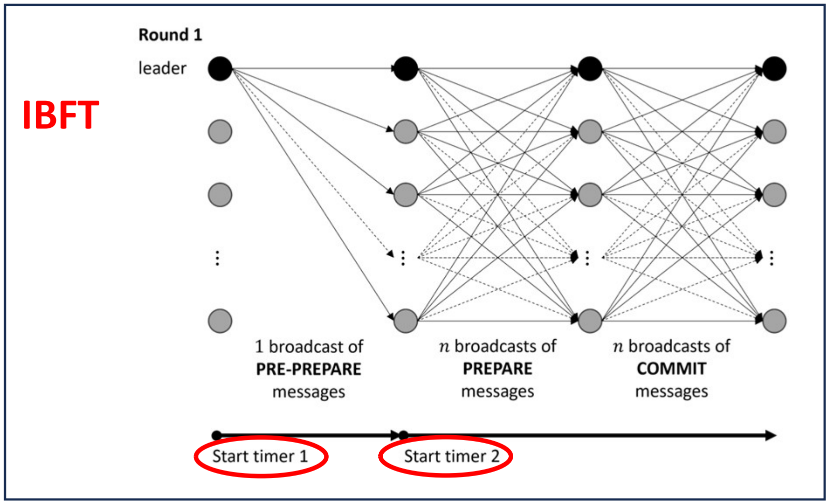
The Istanbul BFT Consensus Algorithm

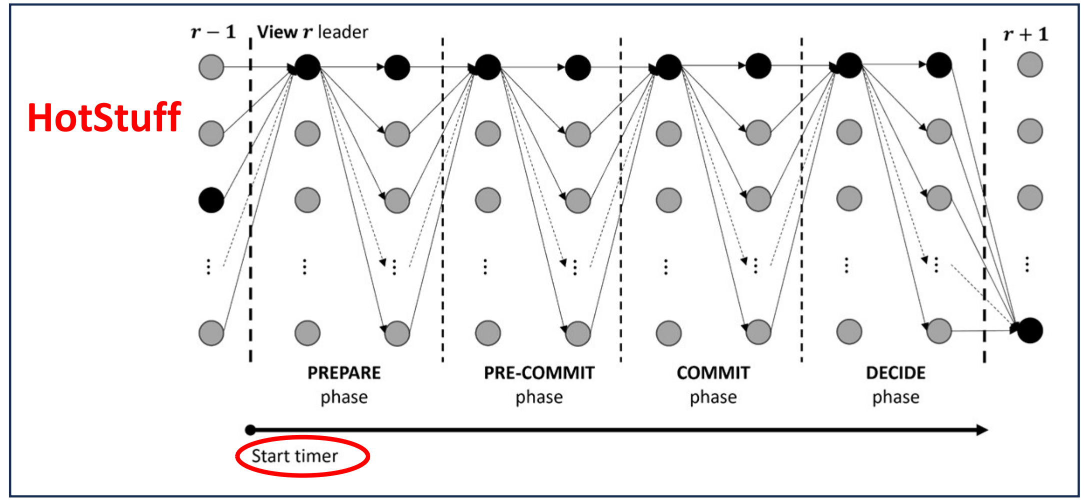
Henrique Moniz
Quorum Engineering

May 20, 2020

Abstract

This paper presents IBFT, a simple and elegant Byzantine fault-tolerant consensus algorithm that is used to implement state machine replication in the *Quorum* blockchain. IBFT assumes a partially synchronous communication model, where safety does not depend on any timing assumptions and only liveness depends on periods of synchrony.


IBFT
Round 1
leader
1 broadcast of
PRE-PREPARE
messages
n broadcasts of
PREPARE
messages
n broadcasts of
COMMIT
messages
Start timer 1
Start timer 2


HotStuff
$r-1$
View r leader
$r+1$
PREPARE
phase
PRE-COMMIT
phase
COMMIT
phase
DECIDE
phase
Start timer

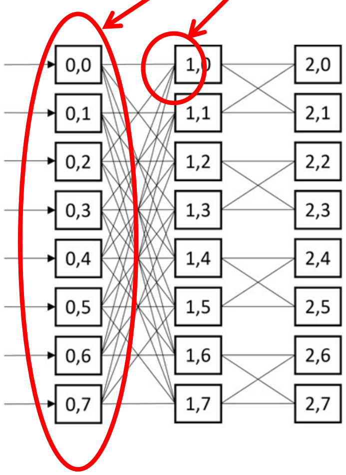
topology

“edge switch” = “switch where validators are attached”

Folded-Clos(8,4)

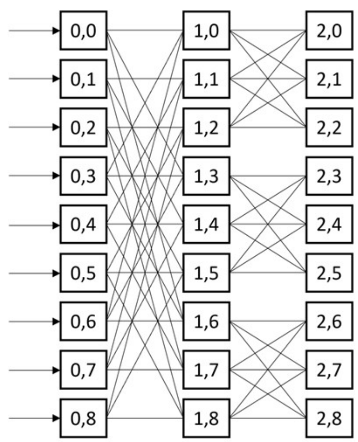
8 edge switches

degree=4



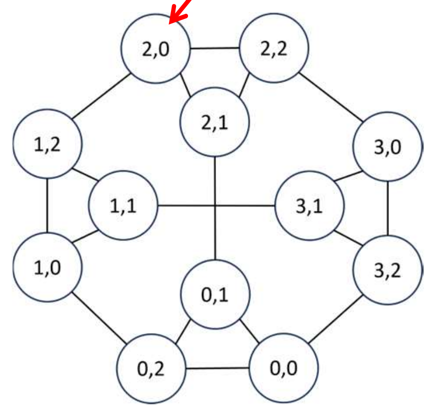
edge switches
at level 1

Folded-Clos(9,3)

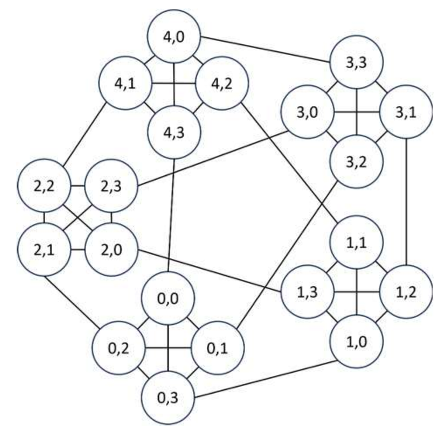


Dragonfly(3)

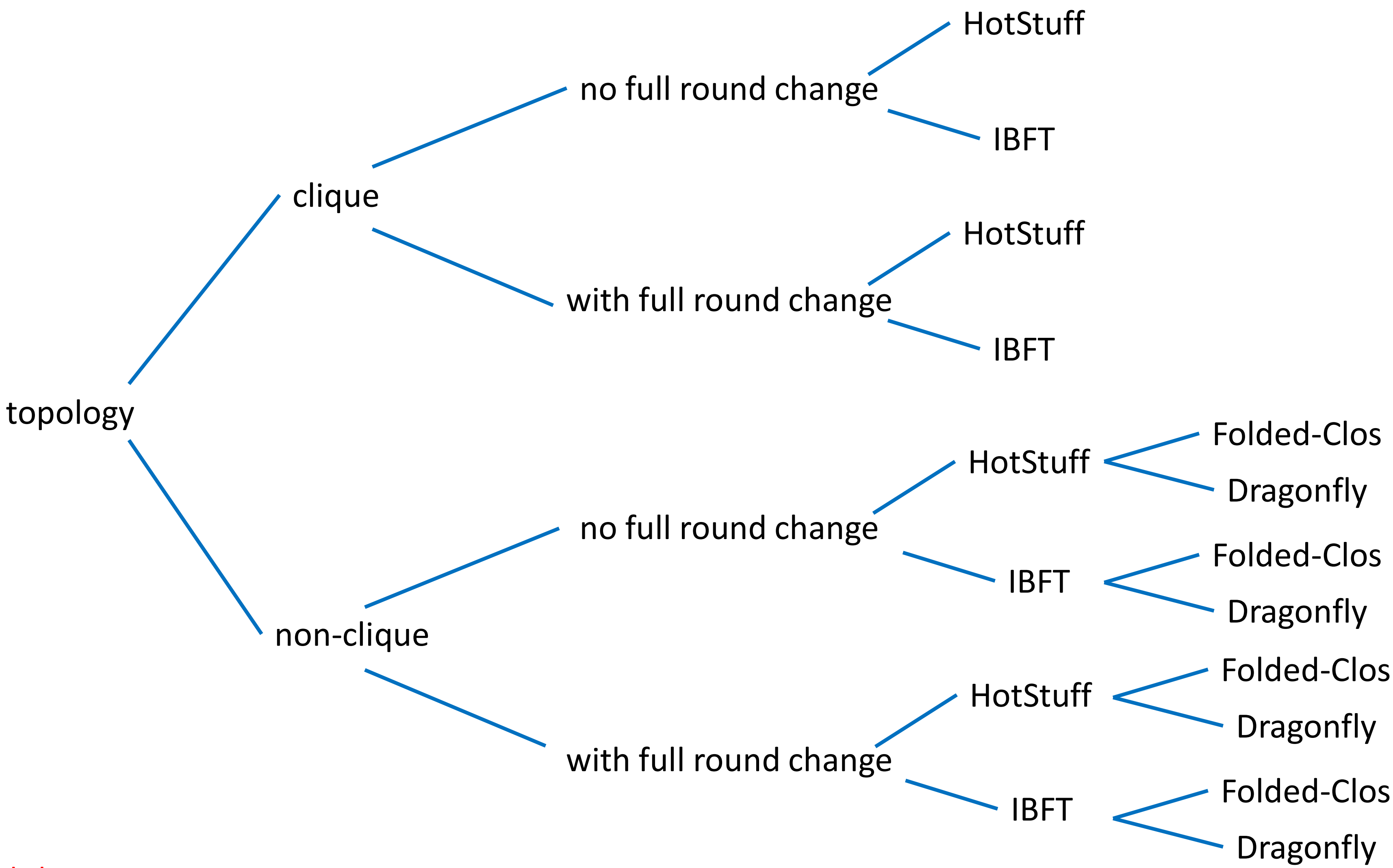
degree=3



Dragonfly(4)



all switches are edge switches



full round change:
all validators undergo round change for the same round (e.g.leader crashes, or validators time out).

system abstraction

application (smart contracts)	
execution (VM/containers)	
data (transactions)	transaction throughput depends on memory pool protocol
consensus (blocks)	HotStuff Istanbul BFT (IBFT)
network (switches)	Clique / Folded-Clos / Dragonfly

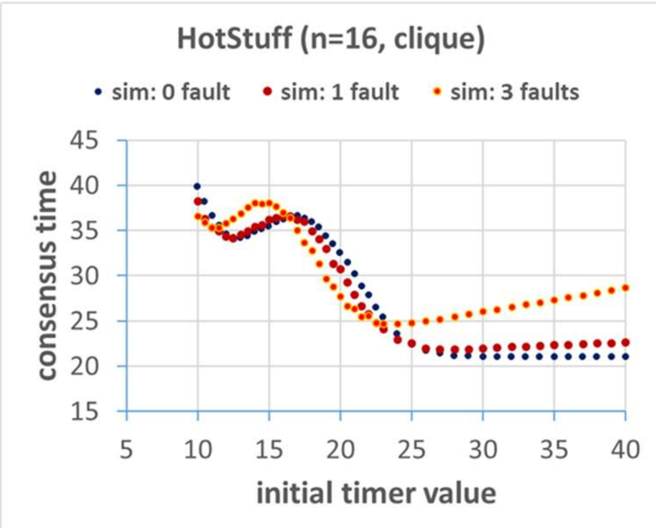
our model

closed model:
every consensus generates a new block

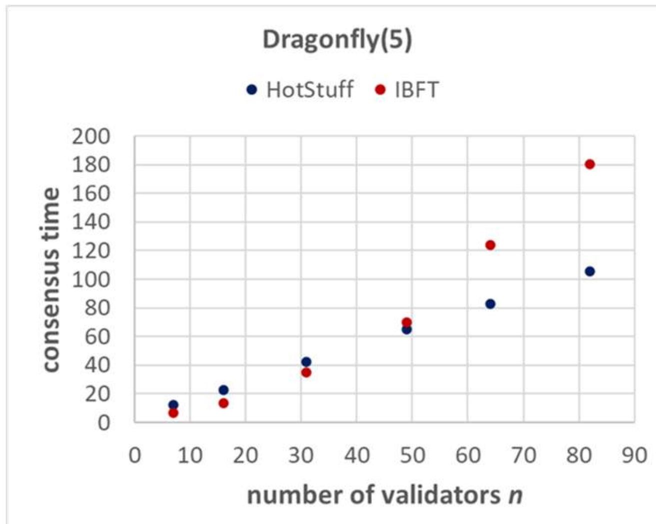
assume crash faults
(arbitrary fault $\Rightarrow$ arbitrary performance)

metric:
average consensus time

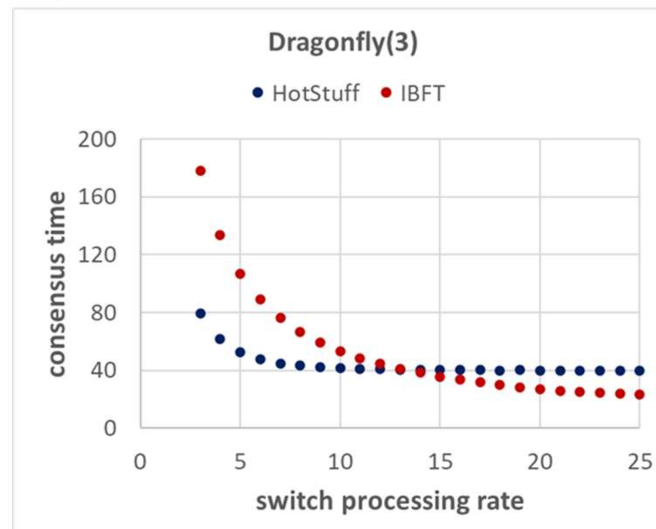
impact of parameters?



impact of scale?



impact of network?



terminology/notation:

n = #validators (participants)

$f = \lfloor (n-1)/3 \rfloor$

n_f = #faults ($n_f \leq f$)

$r = \frac{n_f}{n}$ (fraction of faults)

τ_0 = initial timer value

q = Prob($\underbrace{\text{full round change}}_{\text{change of leader}}$ | nonfaulty leader)

message processing rates

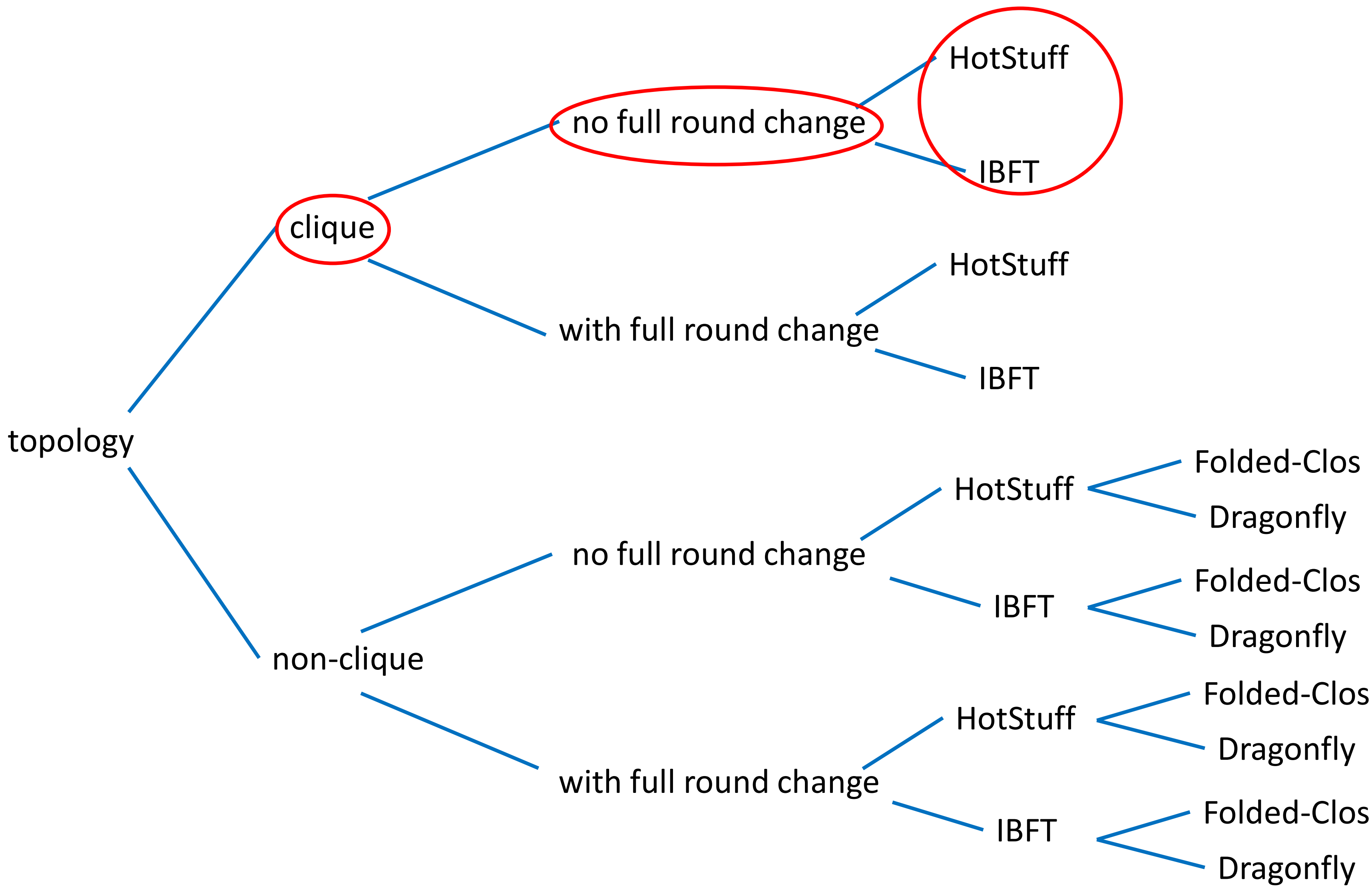
$r_{\mathbf{V}}^{\mathbf{P}}$ at validator

$r_{\mathbf{S}}^{\mathbf{P}}$ at switch

$T_{\mathbf{T}}^{\mathbf{P}}$ = consensus time

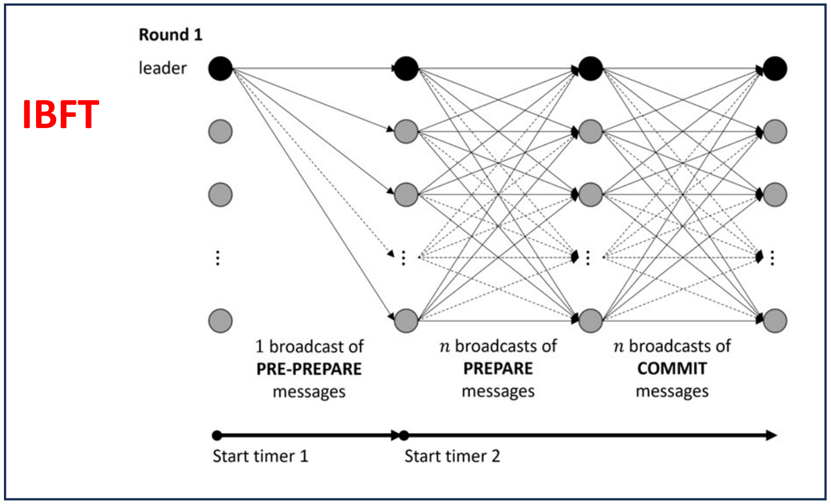
protocol

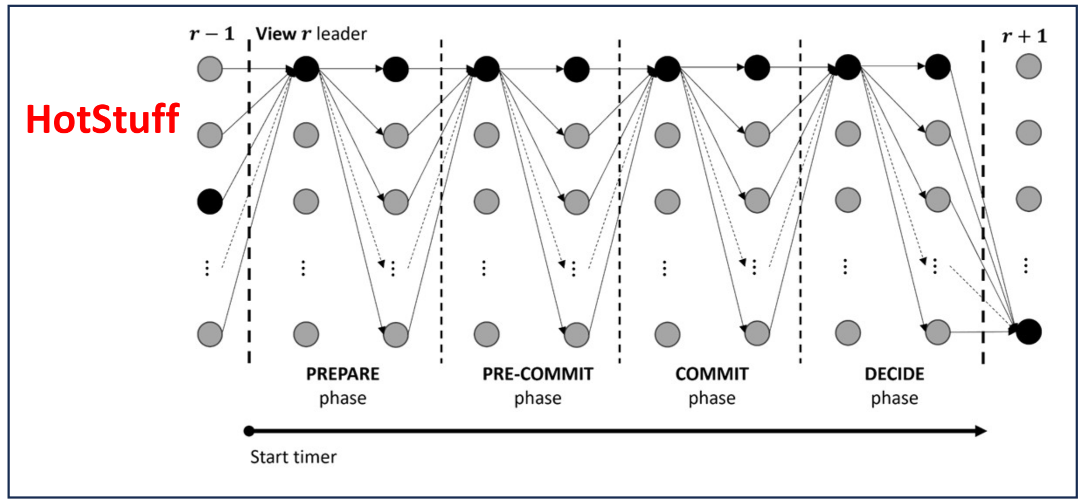
topology


topology
clique
non-clique
no full round change
with full round change
no full round change
with full round change
HotStuff
IBFT
HotStuff
IBFT
HotStuff
IBFT
HotStuff
IBFT
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly

IBFT vs HotStuff (clique; no full round change)

which is faster? scalability?

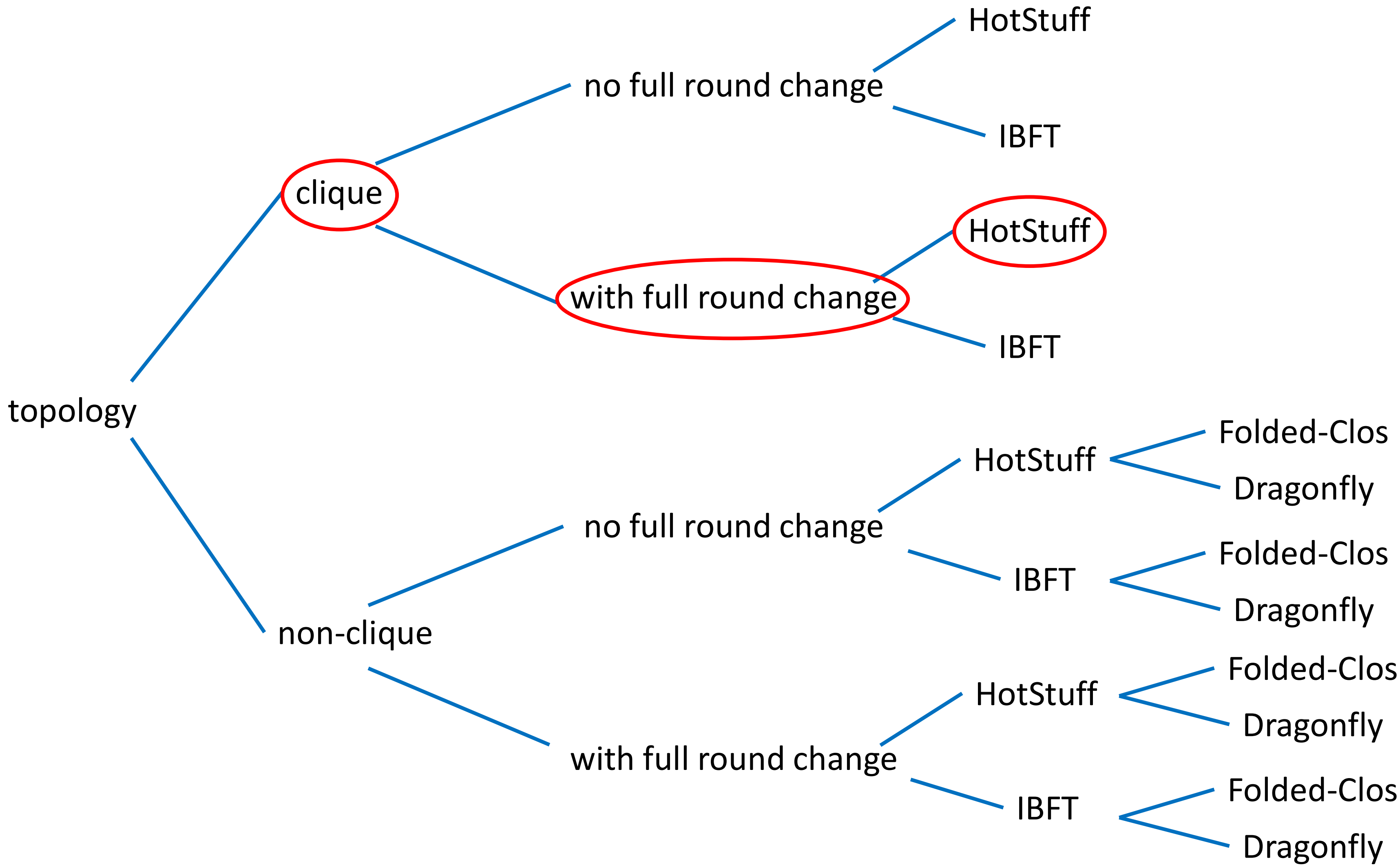




$$ET^{\mathbf{I}}_{\mathbf{C}} = \frac{2n+1}{r^{\mathbf{I}}_{\mathbf{V}}} \approx \frac{2}{r^{\mathbf{I}}_{\mathbf{V}}} n$$

$$ET^{\mathbf{H}}_{\mathbf{C}} = \frac{4n-f+}{r^{\mathbf{H}}_{\mathbf{V}}} \approx \frac{4}{r^{\mathbf{H}}_{\mathbf{V}}} n$$

depends


topology
clique
non-clique
no full round change
with full round change
no full round change
with full round change
HotStuff
IBFT
HotStuff
IBFT
HotStuff
IBFT
HotStuff
IBFT
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly

HotStuff (clique; with full round change)

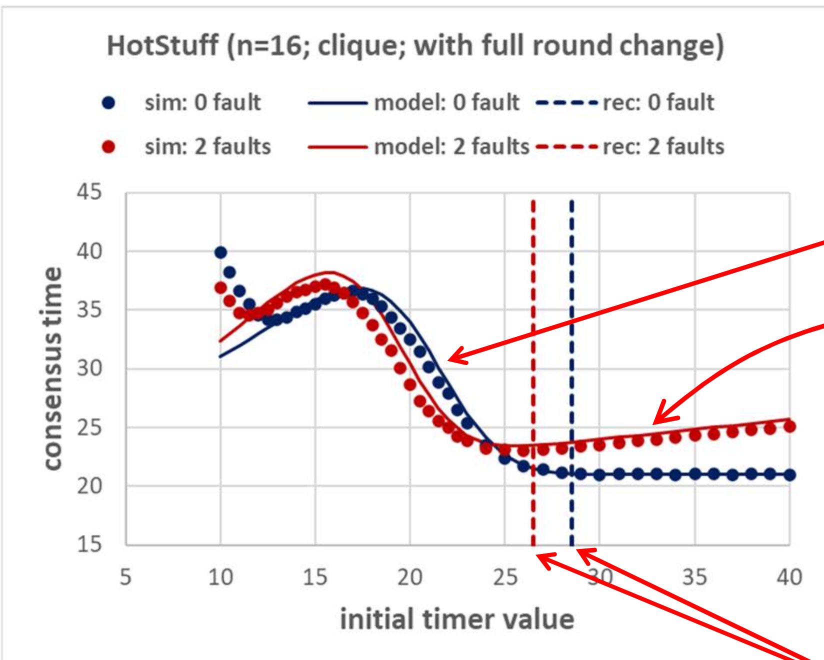
$m^{\mathbf{H}} = 4n - 3n_f - f + 1$ $\quad$ (n_f = #faults)

$$E(T_{\mathbf{C}}^{\mathbf{H}}) = \frac{m^{\mathbf{H}}}{r_{\mathbf{V}}^{\mathbf{H}}} + \frac{r+(1-r)q}{1-2r}\tau_0$$

X_i = validator processing time for i-th message

q = Prob(full round change | nonfaulty leader)

$$= \text{Prob}(\underbrace{X_1 + X_2 + \ldots + X_{m^{\mathbf{H}}}}_{\text{Central Limit Theorem: } N\left(m^{\mathbf{H}}\frac{1}{r_{\mathbf{V}}^{\mathbf{H}}},\ m^{\mathbf{H}}\sigma^2\right)} > \tau_0)$$



0 faults: $E(T_{\mathbf{C}}^{\mathbf{H}}) = \dfrac{m^{\mathbf{H}}}{r_{\mathbf{V}}^{\mathbf{H}}} + q\tau_0$

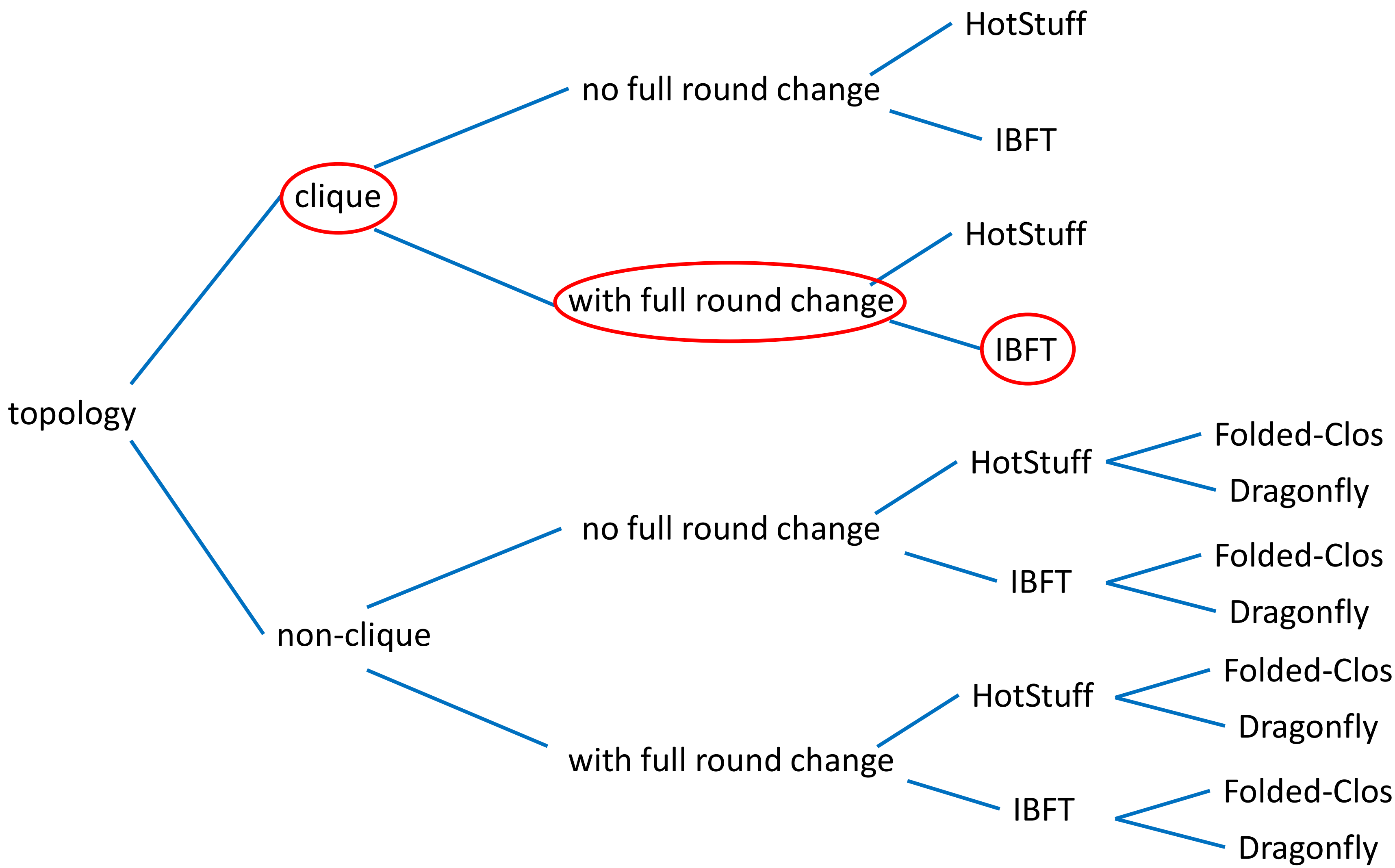
$\left(r = \dfrac{n_f}{n}\right)$

linear

gradient

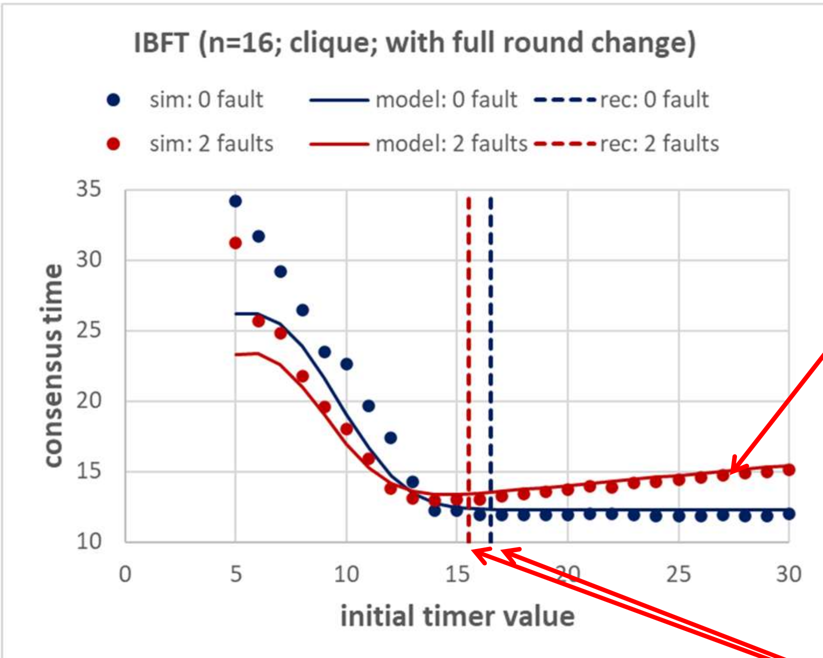
large τ_0 ($q \approx 0$): $E(T_{\mathbf{C}}^{\mathbf{H}}) \approx \dfrac{4n - 3n_f - f + 1}{r_{\mathbf{V}}^{\mathbf{H}}} + \dfrac{r}{1-2r}\tau_0$

recommended $\tau_0^* = m^{\mathbf{H}}\dfrac{1}{r_{\mathbf{V}}^{\mathbf{H}}} + 3\sqrt{m^{\mathbf{H}}}\,\sigma$


topology
clique
non-clique
no full round change
with full round change
HotStuff
IBFT
HotStuff
IBFT
no full round change
with full round change
HotStuff
IBFT
HotStuff
IBFT
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly

IBFT (clique; with full round change)

$$E(T^{\mathbf{I}}_{\mathbf{C}}) = \frac{m^{\mathbf{I}}}{r^{\mathbf{I}}_{\mathbf{V}}} + \frac{r+(1-r)q}{1-2r}\tau_0 + \frac{r+(2\text{-}r)(1\text{-}r)q}{r^{\mathbf{I}}_{\mathbf{V}}}(n\text{-}n_f)$$



q = Prob(full round change | nonfaulty leader)

= Prob($Y > \tau_0$)

$$N\left(m^{\mathbf{I}}\frac{1}{r^{\mathbf{I}}_{\mathbf{V}}},\ m^{\mathbf{I}}\left(1+\underbrace{\left(1+\frac{1}{n_w}\right)}_{\text{added variance from varying validator progress}}\right)\sigma^2\right)$$

$n_w + n - f$

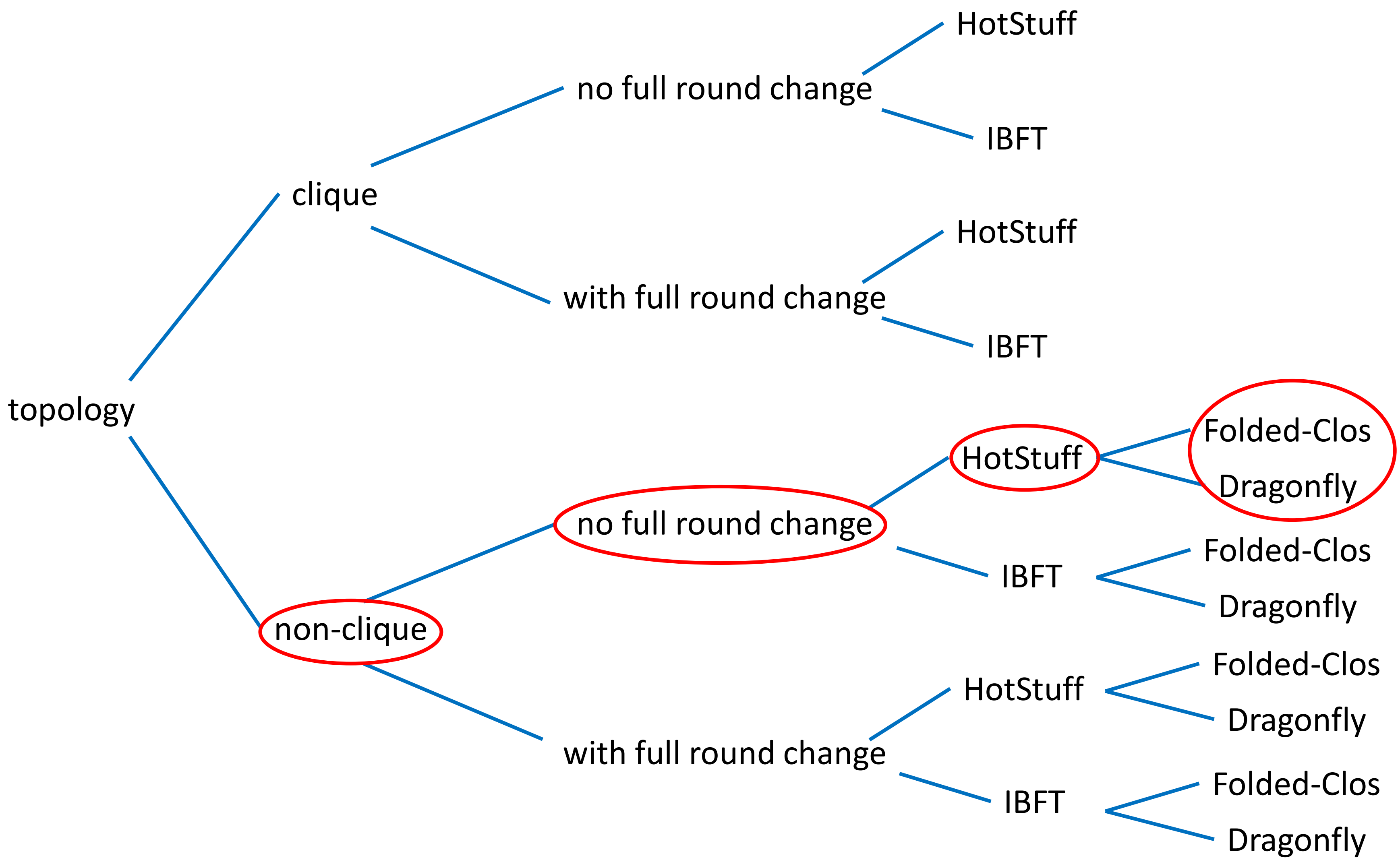
large τ_0 ($q \approx 0$): $\frac{r}{1-2r}\tau_0$ same as HotStuff

HotStuff : $\tau_0^* = m^{\mathbf{H}}\frac{1}{r^{\mathbf{H}}_{\mathbf{V}}} + 3\sqrt{m^{\mathbf{H}}}\,\sigma$

$m^{\mathbf{H}} = 4n - 3n_f - f + 1$

$m^{\mathbf{I}} = 2n - 2n_f - f$

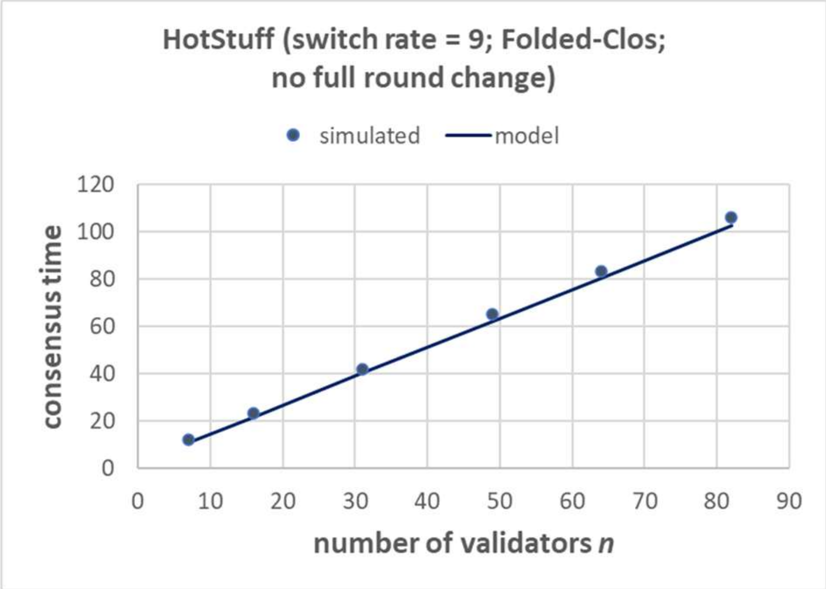
IBFT : $\tau_0^* = m^{\mathbf{I}}\frac{1}{r^{\mathbf{I}}_{\mathbf{V}}} + 3\sqrt{m^{\mathbf{I}}\left(2+\frac{1}{n_w}\right)}\,\sigma$


topology
clique
non-clique
no full round change
with full round change
no full round change
with full round change
HotStuff
IBFT
HotStuff
IBFT
HotStuff
IBFT
HotStuff
IBFT
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly
Folded-Clos
Dragonfly

HotStuff (nonclique; no full round change)

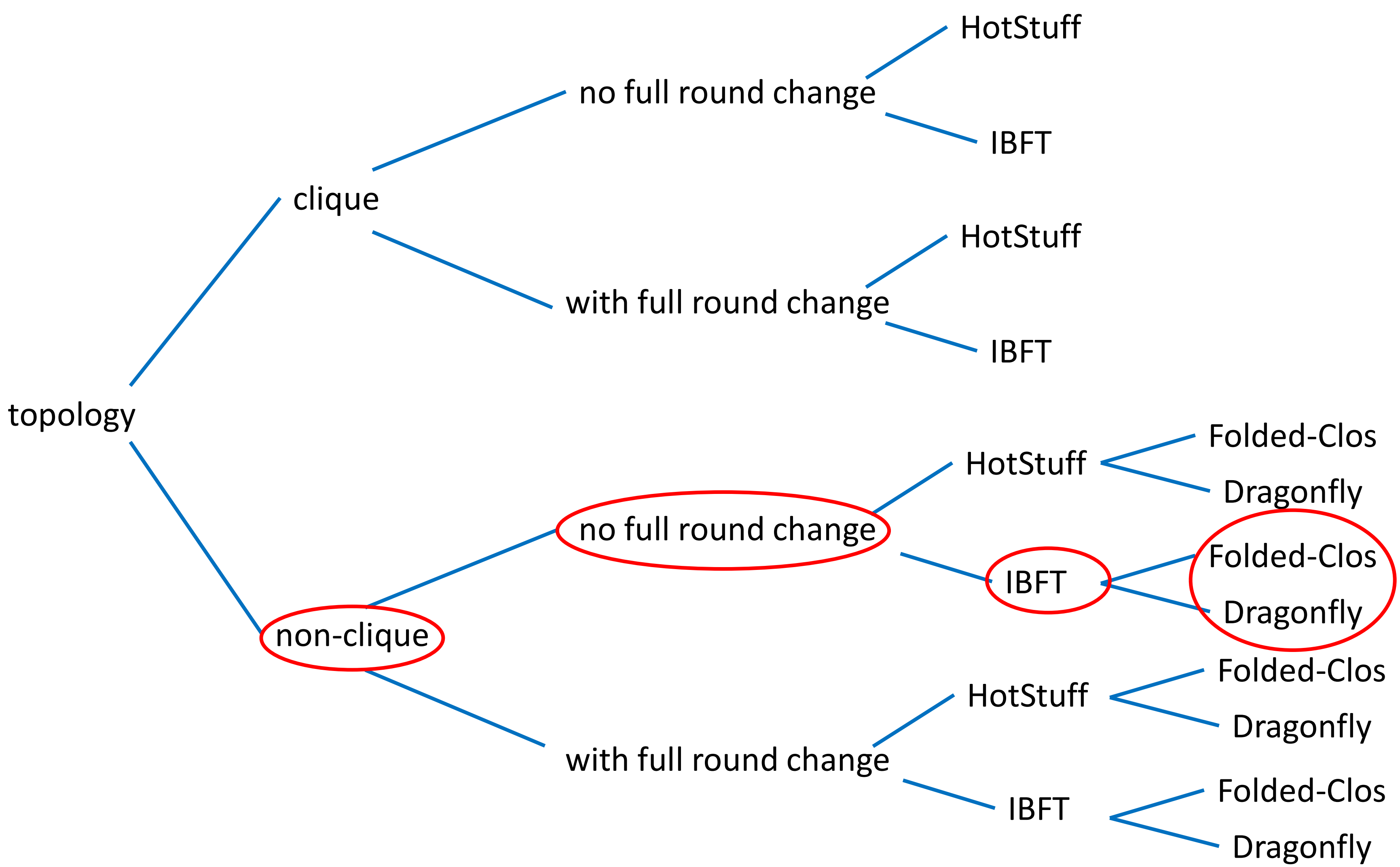
hop count

$$E(T^{\mathbf{H}}_{\mathbf{T}}) = 4\max\{\frac{n\text{-}f\text{-}1}{r^{\mathbf{H}}_{\mathbf{V}}}, \frac{n\text{-}2}{r^{\mathbf{H}}_{\mathbf{S}}}\} + 3\max\{\frac{f+2}{r^{\mathbf{H}}_{\mathbf{V}}}, \frac{n}{r^{\mathbf{H}}_{\mathbf{S}}}\} + 2(\frac{1}{r^{\mathbf{H}}_{\mathbf{V}}} + \frac{h_{\mathbf{T}}}{r^{\mathbf{H}}_{\mathbf{S}}})$$



analysis

- $E(T^{\mathbf{H}}_{\mathbf{T}})$ mostly network delay if $r^{\mathbf{H}}_{\mathbf{S}} < \frac{3}{2} r^{\mathbf{H}}_{\mathbf{V}}$ (independent of n)
- topology has negligible impact if hop count $h_{\mathbf{T}} \ll n$


topology
clique
no full round change
HotStuff
IBFT
with full round change
HotStuff
IBFT
non-clique
no full round change
HotStuff
Folded-Clos
Dragonfly
IBFT
Folded-Clos
Dragonfly
with full round change
HotStuff
Folded-Clos
Dragonfly
IBFT
Folded-Clos
Dragonfly

IBFT vs HotStuff (Dragonfly; no full round change)

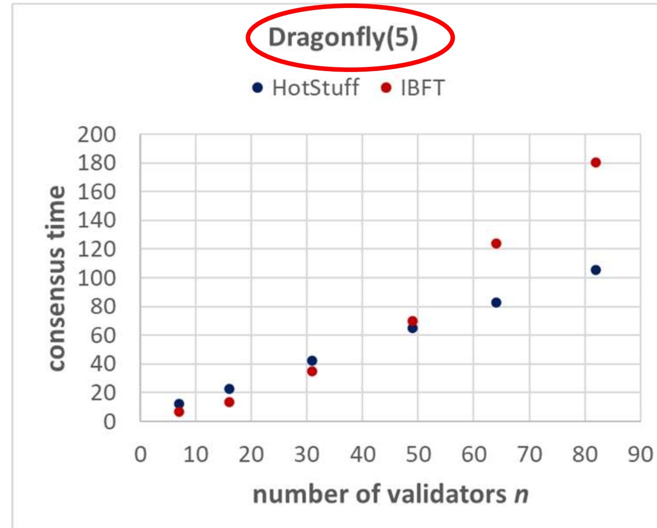
suppose switch rate dominates performance:

HotStuff

$$E(T^{\mathbf{H}}_{\mathbf{T}}) = 4\max\left\{\frac{n\text{-}f\text{-}1}{r^{\mathbf{H}}_{\mathbf{V}}}, \frac{n\text{-}2}{r^{\mathbf{H}}_{\mathbf{S}}}\right\} + 3\max\left\{\frac{f+2}{r^{\mathbf{H}}_{\mathbf{V}}}, \frac{n}{r^{\mathbf{H}}_{\mathbf{S}}}\right\} + 2\left(\frac{1}{r^{\mathbf{H}}_{\mathbf{V}}} + \frac{h_{\mathbf{T}}}{r^{\mathbf{H}}_{\mathbf{S}}}\right)$$

$O(n)$

impact of scale?



IBFT

$$E(T^{\mathbf{I}}_{\mathbf{D}}) = \max\left\{\frac{m^{\mathbf{I}}_{\mathbf{V}}}{r^{\mathbf{I}}_{\mathbf{V}}}, \frac{m^{\mathbf{I}}_{\mathbf{D}}}{r^{\mathbf{I}}_{\mathbf{S}}}\right\} \quad m^{\mathbf{I}}_{\mathbf{D}} = \underbrace{2\left(\frac{n}{c}\right)\left(2(n\text{-}1) - \frac{n}{c}\left(\frac{n}{c} - 1\right) + 2\left(\frac{n}{c}\right)d(d\text{-}1)\right) + (n\text{-}1)}_{O(n^2)}$$

#edge-switches c = #switches

suppose we scale #switches by $c = \frac{n}{k}$, k constant

$$m^{\mathbf{I}}_{\mathbf{D}} = \underbrace{2k\left(2(n\text{-}1) - k(k - 1) + 2n\frac{d(d-1)}{d(d+1)}\right) + (n\text{-}1)}_{O(n)}$$

IBFT vs HotStuff (nonclique; no full round change)

case: HotStuff dominated by computation time and
IBFT dominated by switching time

what determines
the crossover point?

$$E(T_{\mathbf{D}}^{\mathbf{H}}) = 4\max\left\{\frac{n\text{-}f\text{-}1}{r_{\mathbf{V}}^{\mathbf{H}}}, \frac{n\text{-}2}{r_{\mathbf{S}}^{\mathbf{H}}}\right\} + 3\max\left\{\frac{f+2}{r_{\mathbf{V}}^{\mathbf{H}}}, \frac{n}{r_{\mathbf{S}}^{\mathbf{H}}}\right\} + 2\left(\frac{1}{r_{\mathbf{V}}^{\mathbf{H}}} + \frac{h_{\mathbf{D}}}{r_{\mathbf{S}}^{\mathbf{H}}}\right)$$

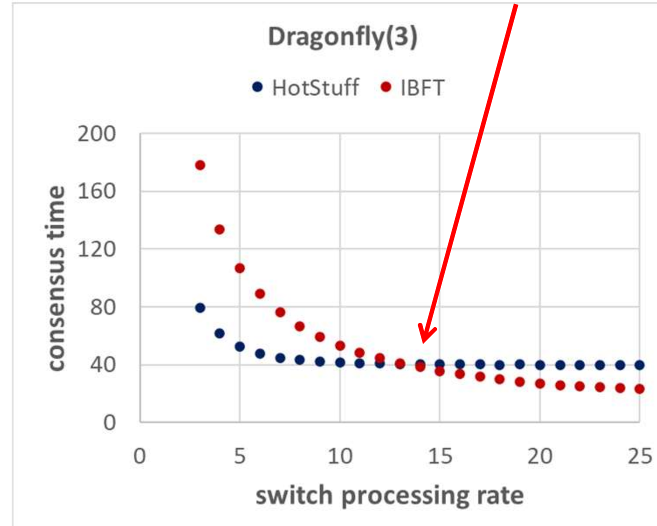
$$= 4\,\frac{n\text{-}f\text{-}1}{r_{\mathbf{V}}^{\mathbf{H}}} + 3\,\frac{f+2}{r_{\mathbf{V}}^{\mathbf{H}}} + 2\frac{1}{r_{\mathbf{V}}^{\mathbf{H}}}$$

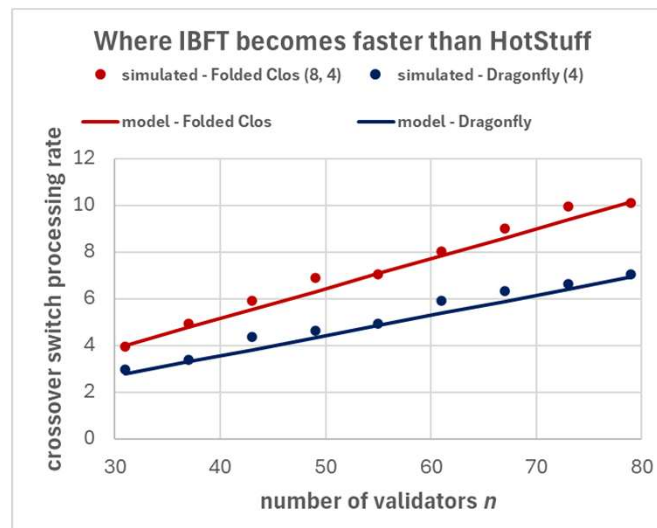
$$E(T_{\mathbf{D}}^{\mathbf{I}}) = \max\left\{\frac{m_{\mathbf{V}}^{\mathbf{I}}}{r_{\mathbf{V}}^{\mathbf{I}}}, \frac{m_{\mathbf{D}}^{\mathbf{I}}}{r_{\mathbf{S}}^{\mathbf{I}}}\right\} = \frac{m_{\mathbf{D}}^{\mathbf{I}}}{r_{\mathbf{S}}^{\mathbf{I}}}$$

$$\Rightarrow \; E(T_{\mathbf{D}}^{\mathbf{H}}) = E(T_{\mathbf{D}}^{\mathbf{I}}) \text{ when } r_{\mathbf{S}}^{\mathbf{I}} \approx \frac{m_{\mathbf{D}}^{\mathbf{I}}}{4n - f + 4}\, r_{\mathbf{V}}^{\mathbf{H}}$$ for Dragonfly

Similarly

$$E(T_{\mathbf{F}}^{\mathbf{H}}) = E(T_{\mathbf{F}}^{\mathbf{I}}) \text{ when } r_{\mathbf{S}}^{\mathbf{I}} \approx \frac{m_{\mathbf{F}}^{\mathbf{I}}}{4n - f + 4}\, r_{\mathbf{V}}^{\mathbf{H}}$$ for Folded-Clos





Folded-Clos vs Dragonfly (**IBFT**; no full round change)

HotStuff

topology has minimal impact
if hopcount << n

$$E(T^{\mathbf{H}}_{\mathbf{T}}) = 4\max\{\frac{n\text{-}f\text{-}1}{r^{\mathbf{H}}_{\mathbf{V}}}, \frac{n\text{-}2}{r^{\mathbf{H}}_{\mathbf{S}}}\} + 3\max\{\frac{f+2}{r^{\mathbf{H}}_{\mathbf{V}}}, \frac{n}{r^{\mathbf{H}}_{\mathbf{S}}}\} + 2(\frac{1}{r^{\mathbf{H}}_{\mathbf{V}}} + \frac{h_{\mathbf{T}}}{r^{\mathbf{H}}_{\mathbf{S}}})$$

⇨ Folded-Clos and Dragonfly have similar consensus time for HotStuff

---

IBFT

$$E(T^{\mathbf{I}}_{\mathbf{T}}) = \max\{\frac{m^{\mathbf{I}}_{\mathbf{V}}}{r^{\mathbf{I}}_{\mathbf{V}}}, \frac{m^{\mathbf{I}}_{\mathbf{T}}}{r^{\mathbf{I}}_{\mathbf{S}}}\}$$

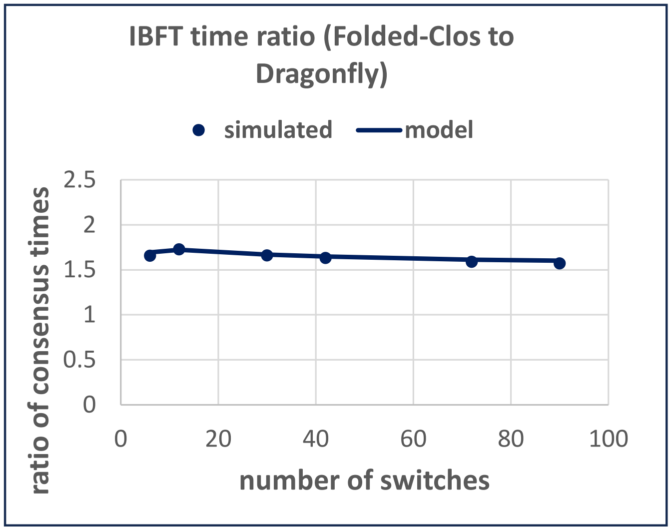
topology has minimal impact if computation time dominates

if communication time dominates (k_F, k_D = #validators per edge-switch):

$$\frac{E(T^{\mathbf{I}}_{\mathbf{F}})}{E(T^{\mathbf{I}}_{\mathbf{D}})} = \frac{m^{\mathbf{I}}_{\mathbf{F}}}{m^{\mathbf{I}}_{\mathbf{D}}} = \frac{2k_F(2(n-1) - k_F(k_F-1)) + (n-1)}{2k_D(2(n-1) - k_D(k_D-1) + 2k_D d(d-1)) + (n-1)} \approx \frac{12}{8+(1/k_D)}$$

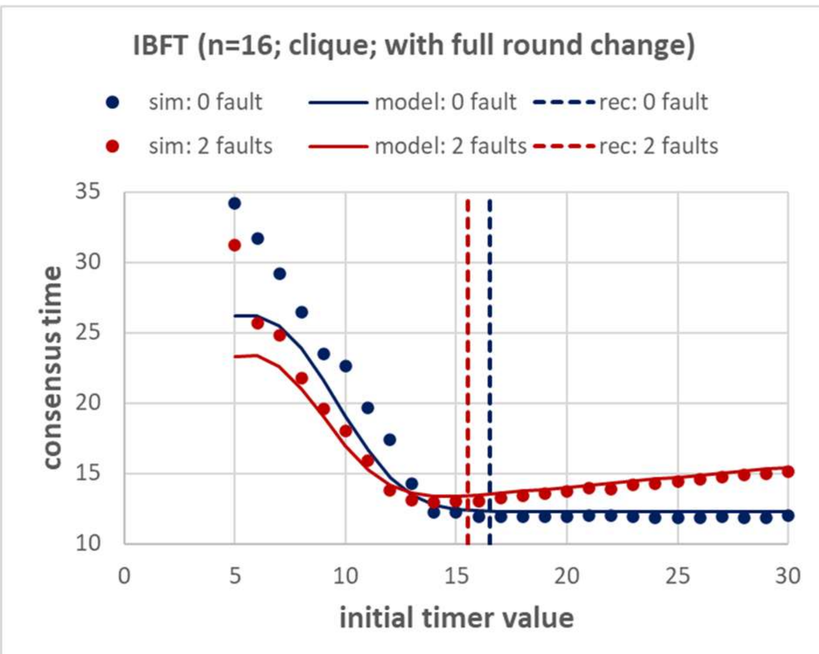
for large n and the same #switches

⇨ for fixed k, Folded-Clos is consistently slower than Dragonfly



The impact of crash tolerance on consensus time for blockchains

- faults (n_f) affect consensus time non-monotonically

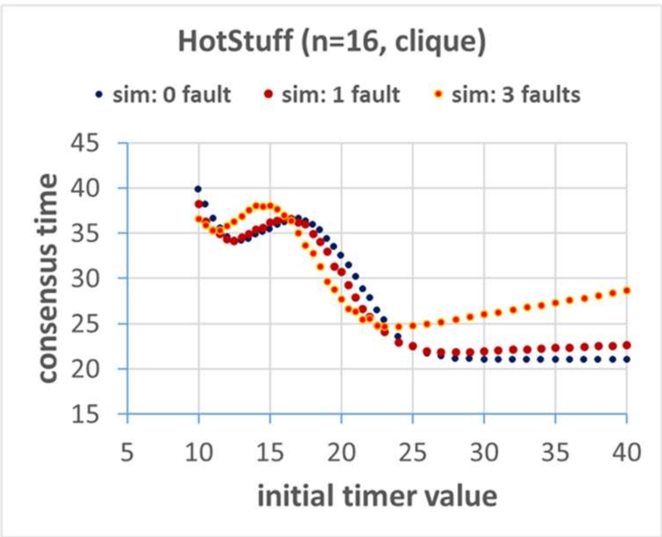


$$E(T^{\mathbf{I}}_{\mathbf{C}}) = \frac{m^{\mathbf{I}}}{r^{\mathbf{I}}_{\mathbf{V}}} + \frac{r+(1-r)q}{1-2r}\tau_0 + \frac{r+(2-r)(1-r)q}{r^{\mathbf{I}}_{\mathbf{V}}}(n-n_f)$$

The impact of crash tolerance on consensus time for blockchains

- faults (n_f) affect consensus time non-monotonically
- fault tolerance (τ_0) affects consensus time non-monotonically

impact of parameters?

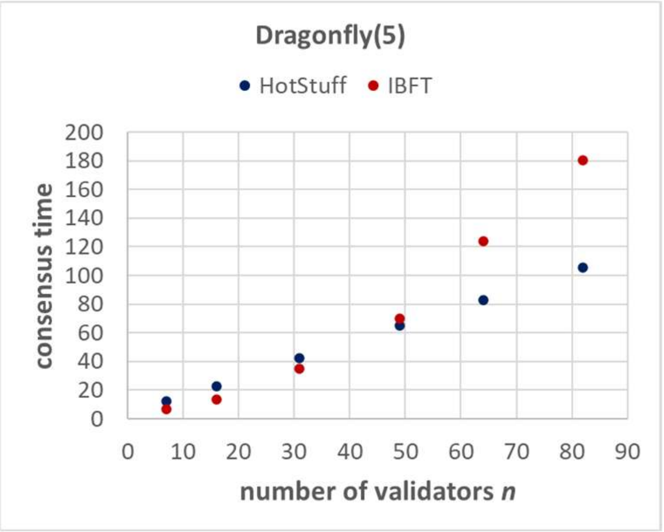


$$E(T_{\mathbf{C}}^{\mathbf{H}}) = \frac{m^{\mathbf{H}}}{r_{\mathbf{V}}^{\mathbf{H}}} + \frac{r+(1-r)q}{1-2r}\tau_0$$

The impact of crash tolerance on consensus time for blockchains

- faults (n_f) affect consensus time non-monotonically
- fault tolerance (τ_0) affects consensus time non-monotonically
- which protocol scales better?

impact of scale?



suppose we scale #switches by $c = \frac{n}{k}$, k constant

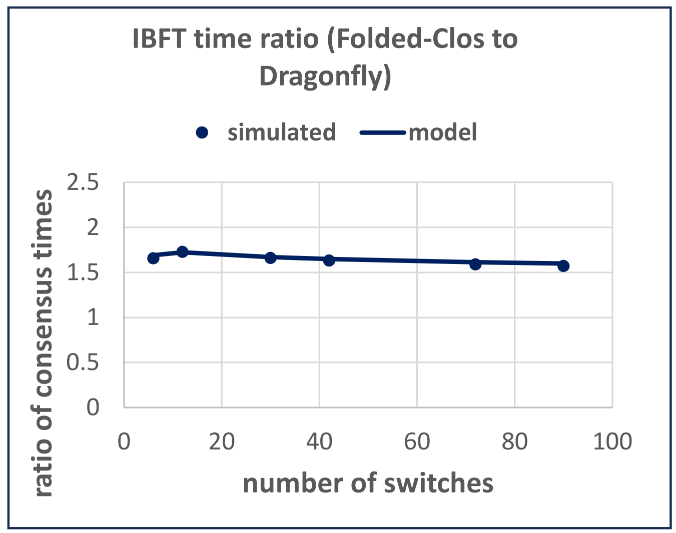
$$m^{\mathbf{I}}_{\mathbf{D}} = \underbrace{2k\,(2(n\text{-}1) - k(k - 1) + 2n\frac{d(d-1)}{d(d+1)}) + (n\text{-}1)}_{O(n)}$$

The impact of crash tolerance on consensus time for blockchains

- faults (n_f) affect consensus time non-monotonically
- fault tolerance (τ_0) affects consensus time non-monotonically
- which protocol scales better? depends on how the network scales

The impact of crash tolerance on consensus time for blockchains

- faults (n_f) affect consensus time non-monotonically
- fault tolerance (τ_0) affects consensus time non-monotonically
- which protocol scales better? depends on how the network scales
- which topology is faster?



$$\frac{E(T_{\mathbf{F}}^{\mathbf{I}})}{E(T_{\mathbf{D}}^{\mathbf{I}})} \approx \frac{12}{8+(1/k_D)}$$

#validators/edge-switch

The impact of crash tolerance on consensus time for blockchains

- faults (n_f) affect consensus time non-monotonically
- fault tolerance (τ_0) affects consensus time non-monotonically
- which protocol scales better? depends on how the network scales
- which topology is faster for IBFT? depends on #validators/edge-switch

The impact of crash tolerance on consensus time for blockchains

Y.C. Tay

(keynote for *IWSF&SHIFT 2025*, Sao Paolo, Brazil)