Mechanical Proofs of Fault-tolerant Clock Synchronization

N. Shankar

Computer Science Laboratory
SRI International

Overview

Introduction to clock synchronization protocols?

A schematic formulation of clock synchronization (Schneider).

The Interactive Convergence Algorithm (Lamport/Melliar-Smith).

Verification of Schneider's formulation (Shankar).

Verification of Interactive Convergence (Rushby/von Henke).

A hardware-oriented clock synchronization protocol (Infis/Moore).

Verification of Infis/Moore's protocol (Rushby/Shankar).

The EHDM Specification/Verification Environment.

Conclusions.

Main Observations

- Fault-tolerant clock synchronization is a critical component of a real-time control system.
- Proofs of the correctness of clock synchronization are complex and subtle.
- Informal proofs tend to be tenuous in these domains.
- Formal verification is a useful way to reduce errors and achieve reliable designs.
- Specification/Verification could contribute to the scientific foundations of reliable engineering.

Fault-tolerant systems

- Critical real-time control systems such as "fly-by-wire" digital avionics.
- Replicated processors are used to provide hardware fault-tolerance.
- Results are periodically voted.
- Clocks must be synchronized to ensure approximately synchronous behaviour across nonfaulty processors.

Clock Synchronization

- Clocks start synchronized.
- Over time, the clocks drift apart.
- The clocks are periodically synchronized by
  - an exchange of clock values
  - computation of a mutually agreeable clock value
  - adjustment of the logical clock

Byzantine Clocks

Three clocks A, B, C.

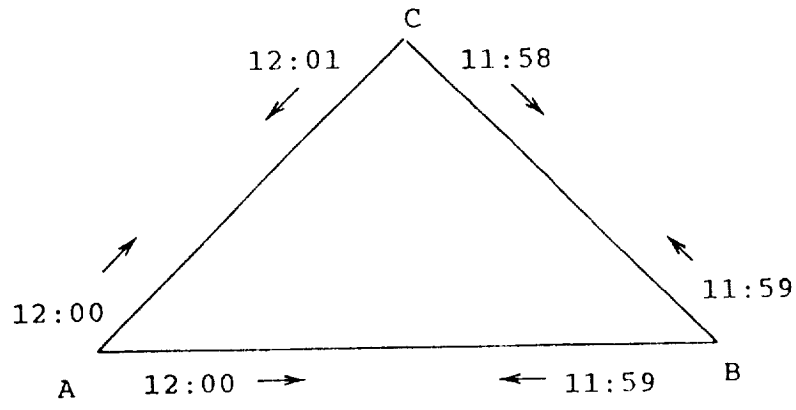
Suppose clocks drift away from real time by upto a minute an hour.

C is faulty.

Clocks resynchronize around noon and exchange clock values.

A reads $12:00$ and B reads $11:59$

A transmits $12:00$ to B and C.
B transmits $11:59$ to A and C.
C maliciously transmits $12:01$ to A; $11:58$ to B.



Byzantine Clocks

Three clocks A, B, C.

Clocks drift from real time by upto a minute an hour.

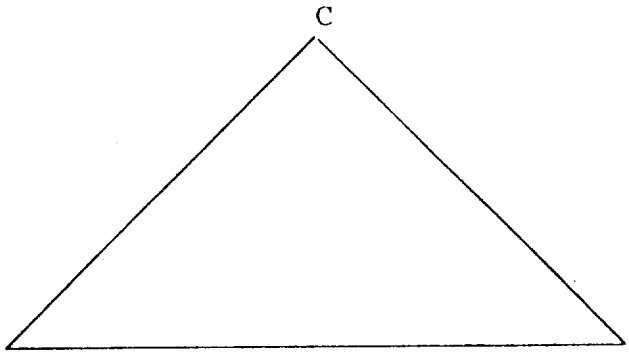
C is faulty.

Clocks resynchronize around noon and exchange clock values.

A reads $12:00$ and B reads $11:59$

A resets its clock to the mean of the acceptable clock values, i.e., $12:00$.
B similarly resets itself to $11:59$.

A and B are not any closer following resynchronization.


C
A
B
12:00
11:59

Clock Generalities

No global clocks — single point of failure, therefore not fault-tolerant.

Synchronization is with respect to other clocks, not *real time*, though such protocols do exist.

Clocks drift at rate ρ with respect to real time.

Period of drift R between resynchronization rounds.

ϵ bounds the error in reading clock values.

To keep clocks synchronized to within δ, clocks should be within δ_s following resynchronization, and

$$\delta > \delta_s + 2\rho R$$

Each clock uses the same *convergence function* to synchronize to within δ_s.

Typical numbers (from Rushby/von Henke)

Parameter	Value	Explanation
N	6	No. of Clocks
R	104.8 msec.	Period
δ_0	132 μsec.	Initial skew
ϵ	66.1 μsec.	Reading error
ρ	15×10^{-6}	Drift rate
δ	271 μsec. $(F = 1)$	Maximum skew

Clock Requirements

- R1: At any instant, two nonfaulty clock readings should be no further than δ apart.

- R2: There should be a small bound on the adjustment needed to resynchronize a clock.

Schneider's Schema

A generalization of various protocols consisting of:

- Assumptions on the behavior of nonfaulty physical clocks.
- Constraints on the computation of nonfaulty logical clocks.

These assumptions and constraints are used to derive a bound on the *skew* between two nonfaulty logical clocks, i.e.

$$|LC_p(t) - LC_q(t)| \leq \delta$$

Physical Clock Assumptions

N clocks with at most F faulty.

t_p^i is the time at which p resets its clock for the i'th time.

Interval between resets is bounded:

$$r_{min} \leq t_p^{i+1} - t_p^i \leq r_{max}$$

Skew between resets is bounded: $|t_p^i - t_q^i| \leq \beta$

Bounded drift rate w.r.t. real time: for $s > t$

$$(s-t)(1-\rho) \leq C_p(s) - C_p(t) \leq (s-t)(1+\rho)$$

Logical Clock Assumptions

A Convergence function Cfn is used to compute the adjusted logical clock.

Let $\Theta_p^i(q)$ be p's reading (estimate) of q's logical clock at time t_p^i.

Then $LC_p(t_p^i) = Cfn(p, \Theta_p^i)$

The i'th adjustment to be applied to the physical clock to derive the logical clock is

$$Adj_p^i = Cfn(p, \Theta_p^i) - C_p(t_p^i)$$

In general the logical clock is defined to be

$$LC_p(t) = C_p(t) + Adj_p^i$$

for $t_p^i \leq t < t_p^{i+1}$

ϵ bounds error with which clocks are read.

Additionally, certain assumptions on behavior of a satisfactory convergence function.

Translation Invariance

Adding X to each clock reading, adds X to the value of the convergence function.

For any X and θ mapping clock numbers to clock readings

$$Cfn(p, (\lambda q{:}\theta(q) + X)) = Cfn(p, \theta) + X$$

Translation invariance is used to compare the values of convergence functions at t_p^i and t_q^i.

Precision Enhancement

Formalizes the intuition that

- the closer the good clocks are to each other
- the closer the different readings of the same good clock
- then the closer the resulting convergence function values

Precision Enhancement (contd.)

Given any predicate P on clocks 0 to $N-1$ that holds of at least $N-F$ clocks.

Given p, q, such that $P(p)$ and $P(q)$.

Given θ_p and θ_q such that

- If $P(l)$ and $P(m)$, then $|\theta_p(l) - \theta_p(m)| \leq Y$
- If $P(l)$ and $P(m)$, then $|\theta_q(l) - \theta_q(m)| \leq Y$
- If $P(l)$, then $|\theta_p(l) - \theta_q(l)| \leq X$

Then there exists a bound $\pi(X, Y)$ such that

$$|Cfn(p, \theta_p) - Cfn(q, \theta_q)| \leq \pi(X, Y)$$

Illustrative example to follow.

Accuracy Preservation

Bounds the adjustment away from a good clock reading.

Given any predicate P on clocks 0 to $N-1$ that holds of at least $N-F$ clocks.

Given that P holds of p and q.

Given θ_p such that whenever $P(l)$ and $P(m)$ for any two clocks l and m, then

$$|\theta_p(l) - \theta_p(m)| \leq Z$$

Then

$$|Cfn(p, \theta_p) - \theta_p(q)| \leq \alpha(Z)$$

That is, if the good clock readings are within Z, the adjustment away from a good clock reading is no more than $\alpha(Z)$.

The Final Result: Agreement

- A1: $\beta \leq r_{min}$
  Synchronization rounds are distinct

- A2: $\delta_0 \leq \delta_s$
  Initial skew no greater than skew immediately following synchronization.

- A3: $\delta_s + 2\rho r_{max} \leq \delta$
  Drift between synchronization rounds is below δ.

- A4: $\pi(2\epsilon + 2\rho\beta, \delta_s + 2\rho(r_{max} + \beta) + 2\epsilon) \leq \delta_s$
  Skew between just synchronized clocks below δ_s.

- A5: $\alpha(\delta_s + 2\rho(r_{max} + \beta) + 2\epsilon) \leq \delta$
  Skew between synchronized and yet to be synchronized clocks below δ.

- Conclusion:

$$\begin{aligned} & t \geq 0 \\ \wedge \quad & \texttt{correct}(p,t) \\ \wedge \quad & \texttt{correct}(q,t) \\ \Rightarrow \quad & |LC(p,t) - LC(q,t)| \leq \delta \end{aligned}$$

  Skew between nonfaulty logical clocks bounded by δ.

Verification of Schneider's Schema using EHDM

Proof consists of:

- 30 axioms involving multiplication, division, and clocks.
- 12 definitions
- 95 lemmas.

Proof took about two man-months using EHDM.

Machine verification takes 1000 to 3500 CPU secs on SUNs.

Numerous inaccuracies in Schneider's original presentation were corrected.

The machine proof adds enormous clarity to Schneider's insightful, but imprecise descriptions and definitions.

Instantiation of Schneider's schema in progress.

Lamport/Melliar-Smith's Interactive Convergence (ICA)

$3F + 1$ clocks needed to tolerate F Byzantine faults.

p records (relative discrepancies of) other clock values when its clock reads iR

"Ignores" clock readings further than Δ away.

Adjusts its clock by the 'egocentric' mean of the acceptable clock differences.

Instantiating Schneider's protocol with ICA

Convergence function:

$$ica(p,\theta) = \Sigma_{l=0}^{N-1} \frac{fix_p(\theta(l),\theta)}{N}$$

where

$$fix_p(x,\theta) = \begin{cases} x & \text{if } |x-\theta(p)| \leq \Delta \\ \theta(p) & \text{otherwise} \end{cases}$$

Translation Invariance: Note that

$$fix_p((\lambda l : \theta(l)+t)(q)) = fix_p(\theta(q),\theta)+t$$

Precision Enhancement of ICA

Given that for all correct l, m

- $|\theta_p(l) - \theta_q(l)| \leq X$
- $|\theta_p(l) - \theta_p(m)| \leq Y$
- $|\theta_q(l) - \theta_q(m)| \leq Y$

We have

$$\begin{aligned} & |ica(p,\theta_p) - ica(q,\theta_q)| \\ \leq\ & X + \frac{FY + 2F\Delta}{N} \\ =\ & \pi(X,Y) \end{aligned}$$

X is negligible, but $Y \approx \Delta$, so

$$\pi(X,Y) \approx \frac{3F\Delta}{N}$$

Since $\Delta \geq \delta + \epsilon$, we get $N > 3F + 1$.

Accuracy Preservation of ICA

If nonfaulty clock readings are Z apart, then F faulty clocks can contribute a further skew of $F\Delta/N$ to the egocentric mean.

So

$$\alpha(Z) \leq Z + \frac{F\Delta}{N}$$

Rushby/von Henke's verification of ICA using EHDM

Around 1–2 man month effort

20 modules

1,550 lines of specification

166 proofs

1 hour elapsed to prove them all on Sun 3/75-8

Verification revealed several minor flaws in a five year old journal proof.

Flaws in Lamport/Melliar-Smith

Main induction incorrect (bad approximations)

Proof of Lemma 4 incorrect (bad approximations); also typographical error in statement

Lemma 1 false in absence of additional constraints in A2

Lemma 2 similarly, also typographical error in statement

Lemma 3 similarly, and unnecessarily general

Missing requirement for S2 in Lemmas 1, 3, 4, and (when repaired) 2

Original Constraints on parameters

C1:

C2:

C3: $\Sigma = \Delta$

C4: $\Delta \gtrsim \delta + \epsilon$

C5: $\delta \gtrsim \delta_0 + \rho R$

C6: $\delta \geq 2(\epsilon + \rho S) + \frac{2m\Delta}{n-m} + \frac{n\rho R}{n-m}$

New Constraints on parameters

C1: $R \geq 3S$

C2: $S \geq \Sigma$

C3: $\Sigma \geq \Delta$

C4: $\Delta \geq \delta + \epsilon + \frac{\rho}{2} S$

C5: $\delta \geq \delta_0 + \rho R$

C6:

$$\delta \geq 2(\epsilon + \rho S) + \frac{2m\Delta}{n-m} + \frac{n\rho R}{n-m} + \frac{n\rho\Sigma}{n-m} + \rho\Delta$$

Infis/Moore's economic approach

Tolerates $F < N/2$ *omission* failures for N clocks.

At clock reading iR, p broadcasts a pulse on its private line.

Say p receives and validates $N - f$ pulses

$(N - F)$'th pulse bounded from above and below by a good pulse.

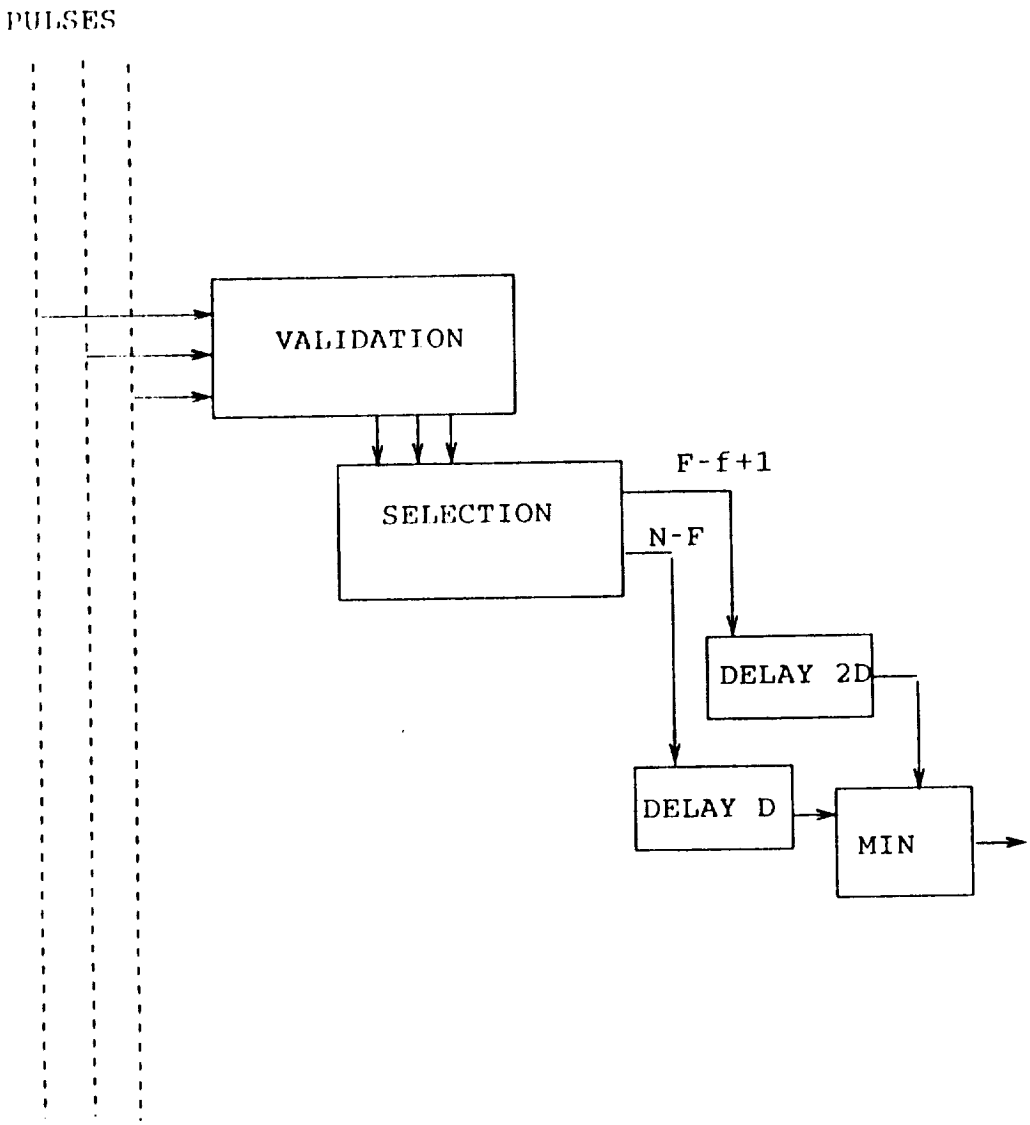
Ditto for $(F - f + 1)$'th pulse.

p starts new clock at earlier of pulse $N - F$ with delay D, or pulse $F - f + 1$ with delay $2D$.

Skew $\delta_s \lesssim D$, and $\delta \lesssim 2D$.

Verification nearly complete using EHDM.
Elaborates significantly on informal proof.

Schemata for Infis/Moore's protocol



Extract from Infis/Moore

(*a*) $T^k_{n-t} \geqslant T_{n-t}$ because the T^k_i are a subset of the T_i

(*b*) $T^k_{n-t} \leqslant T_{n-m}$ because at least one of the times T^k_{n-} ... T^k_{n-f} must be a message from a processor which is actually fault-free (and synchronised) and T_{n-m} is either the time of the message from the last fault-free processor or later

(*c*) $T^k_{n-f} \geqslant T_{n-m}$ because the T_{n-m} is validated by all fault-free processors and must be included in the T^k_i

(*d*) $T^k_{n-f} \leqslant T_{n-g}$ because the T^k_i are a subset of the T_i.

From these inequalities we have that

$$\min \{T_{n-t} + d, T_{n-m}\} \leqslant W \leqslant \min \{T_{n-m} + d, T_{n-g}\} \qquad (1)$$

Now $T^k_{t-f+1} \leqslant T_{n-t}$ for all k and $T^k_{n-f} = T_{n-g}$ for some k, so the validity tests $T^k_{n-f} - T^k_{t-f+1} < 2d$ imply that $T_{n-g} - T_{n-t} < 2d$. Therefore $T_{n-m} - T_{n-t} < d$ or $T_{n-g} - T_{n-m} < d$ (or both).

If $T_{n-m} - T_{n-t} < d$, eqn. 1 reduces to

$$T_{n-m} \leqslant W \leqslant \min \{T_{n-m} + d, T_{n-g}\}$$

implying that W has a range of at most d.

If $T_{n-g} - T_{n-m} < d$, then, using also that $T_{n-g} - T_{n-t} < 2d$, eqn. 1 yields

$$T_{n-g} - d < W \leqslant T_{n-g}$$

implying that W has a range less than d.

Verification of Infis/Moore's protocol

Formalization is fairly close to hardware realization.

Main induction over synchronization rounds completed, as well as all of the important lemmas.

Machine proof is remarkably involved and complex.

Proof took two man-months of effort and covers about 70 dense pages.

Common Errors

Ignoring failures.

Distinguishing real and clock time, and relative versus absolute measurements.

Ignoring small but significant quantities.

Proving one statement but using another.

Imprecise definitions.

Erroneous algebraic manipulations.

Implicit assumptions.

Incorrect assumptions.

Difficulties in verification

Dealing simultaneously with failures, temporal ordering, relative measurements, drift.

Have to be careful not to assume anything about failed clocks.

"Circular definitions" need to be avoided.
E.g., A round ends when various events have taken place.
Various events take place as scheduled if the clock is correct at the end of the round.

Mentally retaining all the relevant facts is difficult.

EHDM specification/verification system

Based on a simply typed higher-order logic with subtyping.

Parametric modules used to structure specifications.

Specifications can be proved to implement other specifications.

Components include parser, typechecker, theorem prover, Hoare sentence prover, and MLS tool.

Theorem prover contains powerful decision procedures for integer and rational inequalities.

New implementation should be ready by end of 1990.

Concluding Observations

Reasoning about fault-tolerant clock synchronization is extremely difficult.

Proofs involve heavy use of inequalities, algebraic manipulations, finite set theory, and induction.

Protocol designers themselves feel the need for mechanized verification tools.

Benefits of such tools are:

- Design discipline
- Efficient location/correction of design errors
- Design library for future reuse
- Standardized language for communicating designs and proofs

Specification and verification technology could contribute effectively to the foundations of reliable engineering.