DSN'01
July 1–4, 2001, Göteborg (Sweden)

How to Model Link Failures: A Perception-based Fault Model

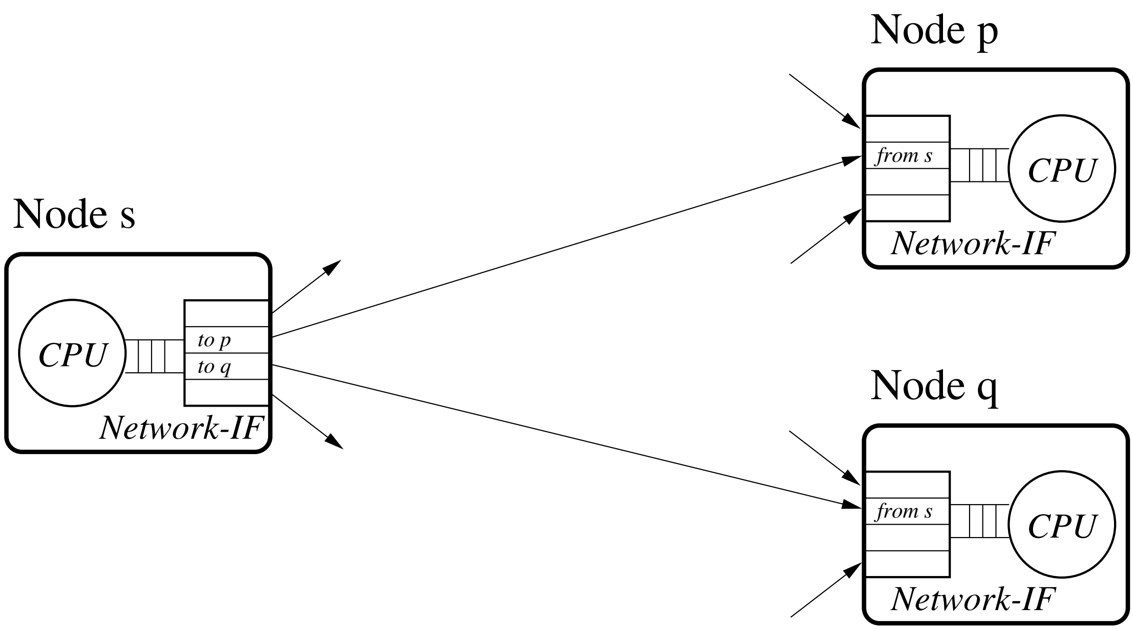
Ulrich Schmid

Technische Universität Wien

Department of Automation
Treitlstraße 1, A-1040 Vienna
s@auto.tuwien.ac.at

June 26, 2001

Motivation (Link Faults)



Many faults $\neq$ sender/receiver CPU exist, like

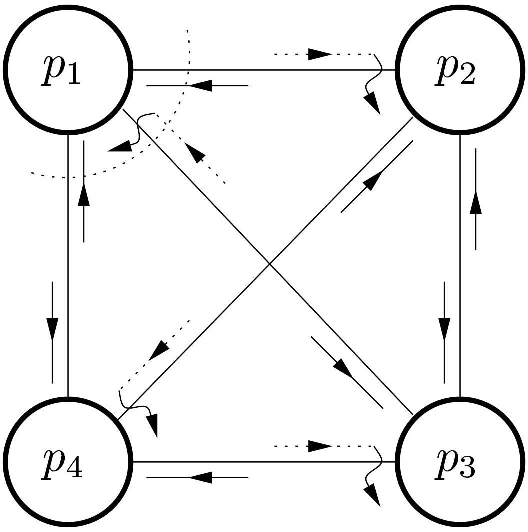
- channel faults
- receiver overruns
- packet preamble synchronization problems
- corrupted data bits

Those **link faults**

- dominate in wireless and high-speed wireline networks
- easily produce asymmetric faults when transmitting time information

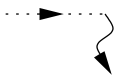
Motivation (Inadequacy Process Fault Model)

Small systems example:



Synchronous system:

- $n = 4$ nodes
- $f_\ell = 1$ link faults per recv. node (in each round)
- systemwide $\leq n \cdot f_\ell = \Omega(n^2)$ link faults



... message loss

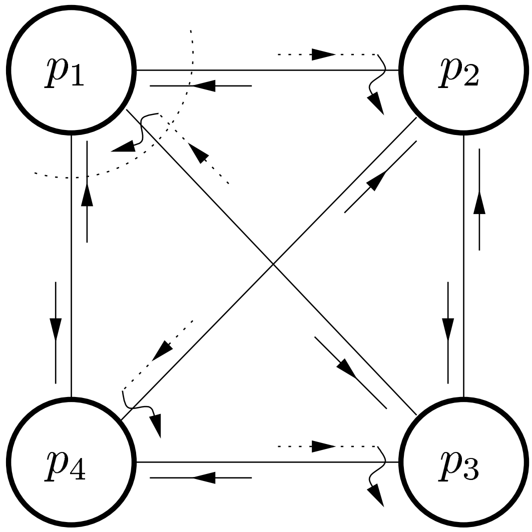
Traditional "global" fault models

- ascribe link faults to [sending] nodes
- $f = 4$ (i.e., all) nodes appear faulty in above example
- $n > 3f + 1$ violated $\Rightarrow$ clock synchronization impossible?

Motivation (Perception-based Fault Model)

Clock synchronization condition just says:

- For any pair of (non-faulty) nodes p, q,
- $|C_p(t) - C_q(t)| \leq \pi$ *independent of other nodes!*



Perception vector $\mathcal{V}_2$:

–	p_2	p_3	p_4

(messages collected by node 2 in a round)

Perception vector $\mathcal{V}_3$:

p_1	p_2	p_3	–

(messages collected by node 3)

Node p's and q's collection of messages from all nodes (perception vectors $\mathcal{V}_p$, $\mathcal{V}_q$)

- differ in at most $2f_\ell = 2$ places

⇒ both contain message from at least 2 (same) senders

⇒ computing new clock values close to each other —and hence clock synchronization— is possible!

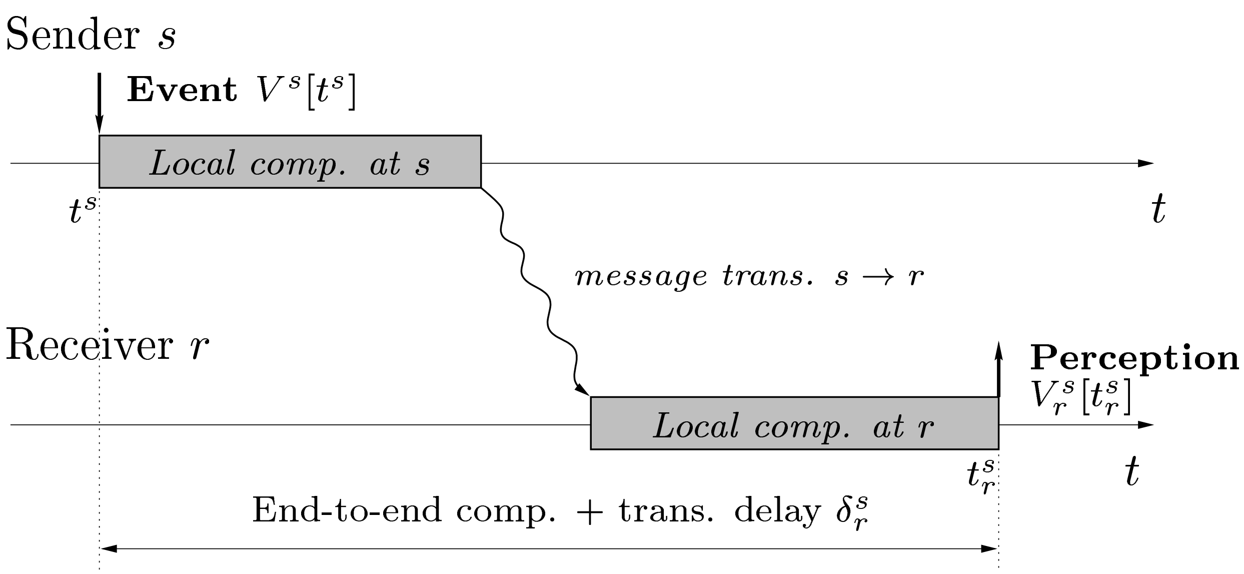
Perception-based Fault Model (Events)

(Virtual) Event $V^s[t^s]$ at node s

- event generation time t^s
- event value V^s
- always considered correct

Node r's perception $V_r^s[t_r^s]$ of $V^s[t^s]$

- result of transmission & local computation at s, r
- occurrence time $t_r^s = t^s + \delta_r^s$
- perceived value V_r^s
- may be both time and value faulty



Perception-based Fault Model (Basic Definition)

Perception vectors

$$\begin{aligned}\mathcal{V}_p(t) &= (V_p^1[t_p^1], V_p^2[t_p^2], \ldots, V_p^n[t_p^n]) \\ \mathcal{V}_q(t) &= (V_q^1[t_q^1], V_q^2[t_q^2], \ldots, V_q^n[t_q^n])\end{aligned}$$

contain all perceptions available at node p, q by time t

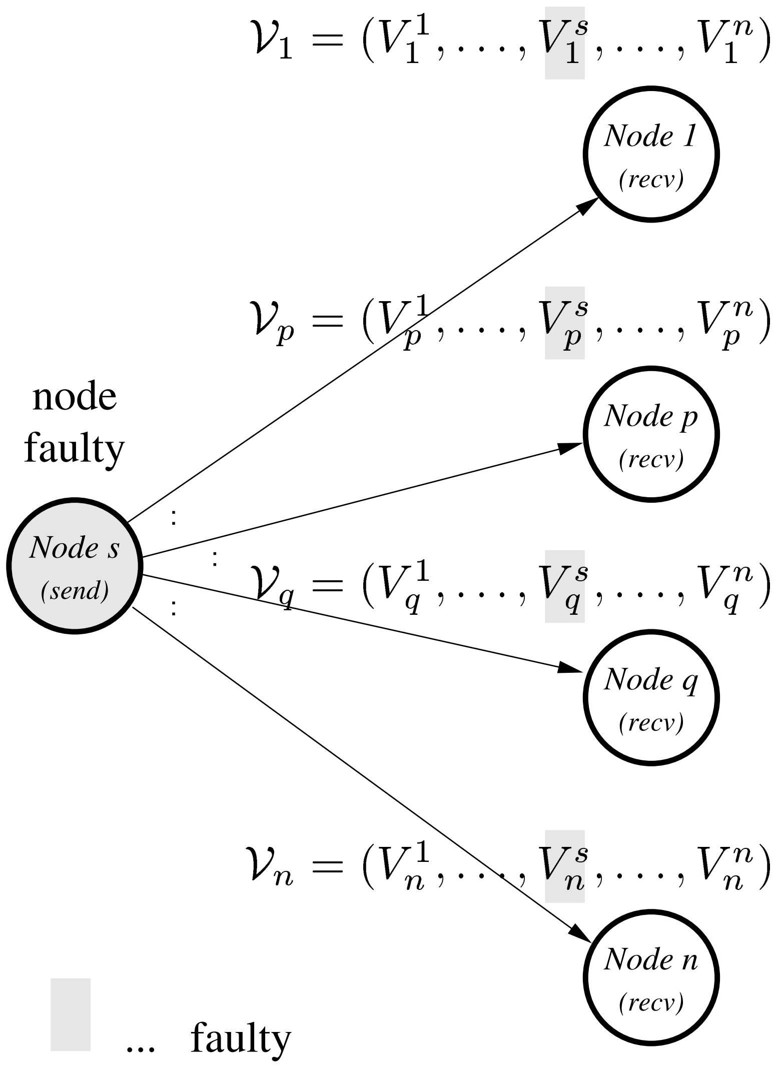
Restrictions on distributed algorithm under consideration ensure

$$\mathcal{V}_q(t) \subseteq \mathcal{V}_q(t') \qquad \text{for } t' \geq t$$

Basic perception-based fault model:

- Define event(s) and perception(s) of interest
- For any pair of perception vectors $\left\{ \begin{array}{c} \mathcal{V}_p(\infty) \\ \mathcal{V}_q(\infty) \end{array} \right\}$,
- consider maximum number f of faulty pairs of perceptions $\left\{ \begin{array}{c} V_p^s \\ V_q^s \end{array} \right\} \in \left\{ \begin{array}{c} \mathcal{V}_p(\infty) \\ \mathcal{V}_q(\infty) \end{array} \right\}$
- possibly distinguish different types of faults (hybrid fault model)

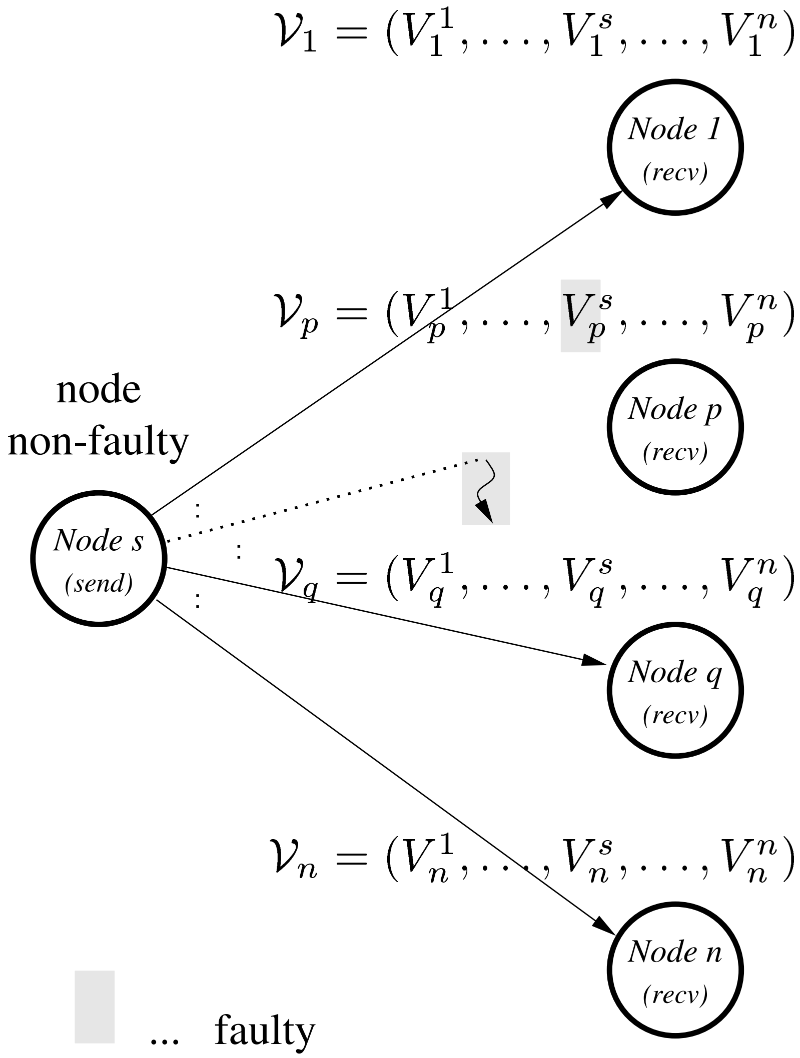
Perception-based Fault Model (Node Faults)



Node fault at node s:

- All perceptions V_r^s (possibly inconsistently) faulty
- Traditional lower bound / impossibility results remain valid

Perception-based Fault Model (Link Faults)



Link fault between sender s and receiver p:

- Only V_p^s in p's perception vector is faulty
- Other (pairs of) perception vectors not affected

$\Rightarrow$ Many $\left(\Omega(n^2)\right)$ link faults tolerable systemwide

Perception-based Fault Model (Value Faults)

Single perception:

- *Omission fault*: $V_p^s = \emptyset$ [and $t_p^s := \infty$]
- *Value fault*: $V_p^s \neq \emptyset$ but $V_p^s \neq V^s$

Pair of perceptions:

- *Symmetric omission fault*: $V_p^s = V_q^s = \emptyset$
- *Symmetric commission fault*: $V_p^s = V_q^s$ is value faulty
- *Asymmetric omission fault*: W.l.o.g. $V_p^s = \emptyset$ and $V_q^s[t_q^s] \neq \emptyset$ is non-faulty
- *Asymmetric commission fault*: W.l.o.g. V_p^s is value faulty and $V_p^s \neq V_q^s$

Perception-based Fault Model (Time Faults)

Clock synchronization:

- Occurence time more important than value
- Uncertainty of $\delta_r^s \Rightarrow$ whole set of admissible values $\boldsymbol{\tau}_p^s$

Single perception:

- *Omission fault*: $t_p^s = \infty$ [and $V_p^s := \emptyset$]
- *Time fault*: $t_p^s < \infty$ but $t_p^s \notin t^s + \boldsymbol{\tau}_p^s$

Pair of perceptions:

- *Symmetric omission fault*: $t_p^s = t_q^s = \infty$
- *Symmetric commission fault*: $t_p^s < \infty$ and/or $t_q^s < \infty$ are/is time faulty and "close" to each other:

$$(t_q^s - \boldsymbol{\tau}_q^s) \cap (t_p^s - \boldsymbol{\tau}_p^s) \neq \emptyset$$

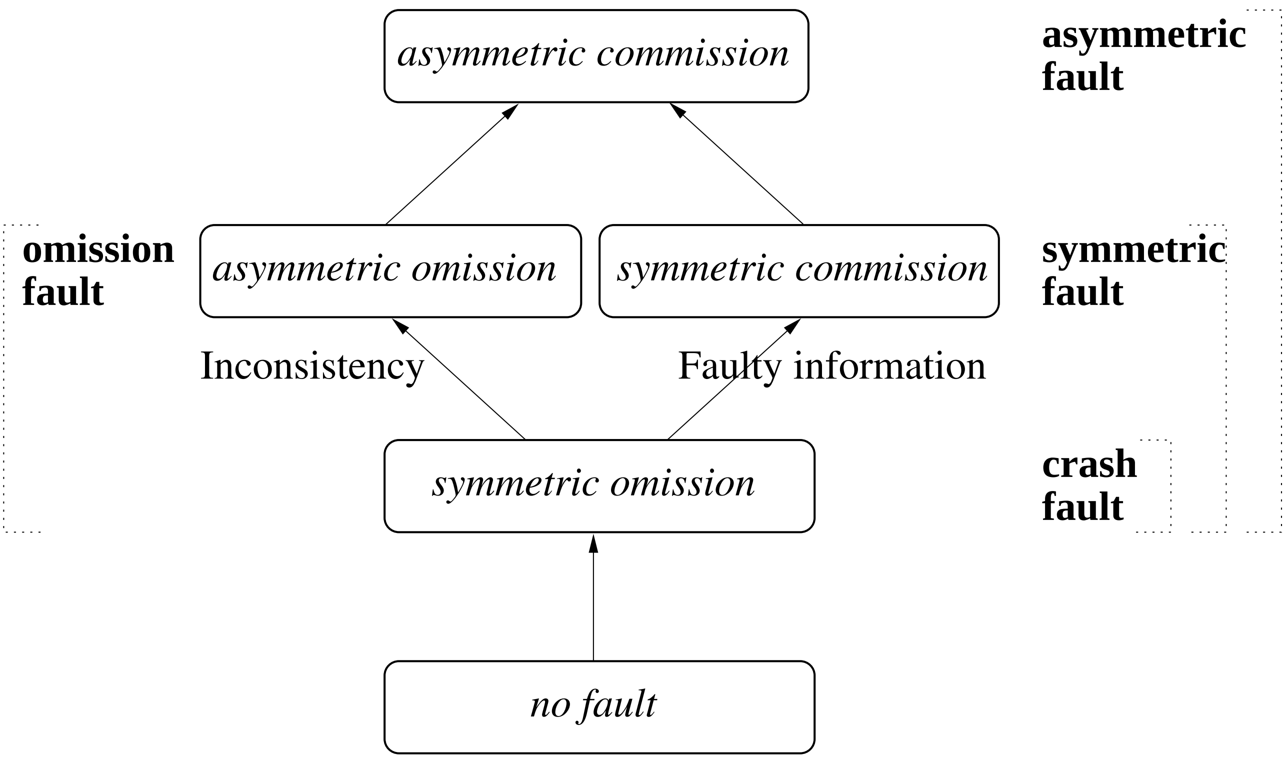
- *Asymmetric omission fault*: W.l.o.g. $t_p^s = \infty$ and $t_q^s < \infty$ is not time faulty
- *Asymmetric commission fault*: W.l.o.g. $t_p^s < \infty$ is time faulty, t_q^s is arbitrary, and

$$(t_q^s - \boldsymbol{\tau}_q^s) \cap (t_p^s - \boldsymbol{\tau}_p^s) = \emptyset$$

Perception-based Fault Model (Fault Classification)

Final fault classification:

- Partially ordered instead of mutual exclusive
- Time and value faults simultaneously



Consistent Broadcasting (Properties)

Implemented via

- broadcast(msg)
- deliver(msg)

atop of unreliable sending/receiving of messages.

Required properties:

P1. *Correctness*: If at least $L > 0$ nodes broadcast msg by time t, then every non-faulty node delivers msg by time $t + t_{del}$

P2. *Unforgeability*: If at most $U \geq 0$ non-faulty nodes broadcast msg by time t, then no non-faulty node delivers msg by t or earlier

P3. *Relay*: If a non-faulty node delivers msg at time t, then every non-faulty node does so by time $t + t_{\Delta}$

Application Consistent Broadcasting

cobegin

    if $C^{k-1}(t) = kP$ /* ready to start C^k */
      $\rightarrow$ broadcast (*round k*);
    fi

    if delivered (*round k*)
      $\rightarrow C^k(t) := kP + \alpha$; /* resynchronize */
    fi

coend

Non-authenticated clock synchronization algorithm of Srikanth & Toueg 1987:

- Exploits tightness of reception by relay property P3
- Achievable precision

$$D_{\max} \geq [(P - \alpha)(1 + \rho) + t_{del}]dr + t_{\Delta}(1 + \rho)$$

with $dr = \rho(2 + \rho)/(1 + \rho)$

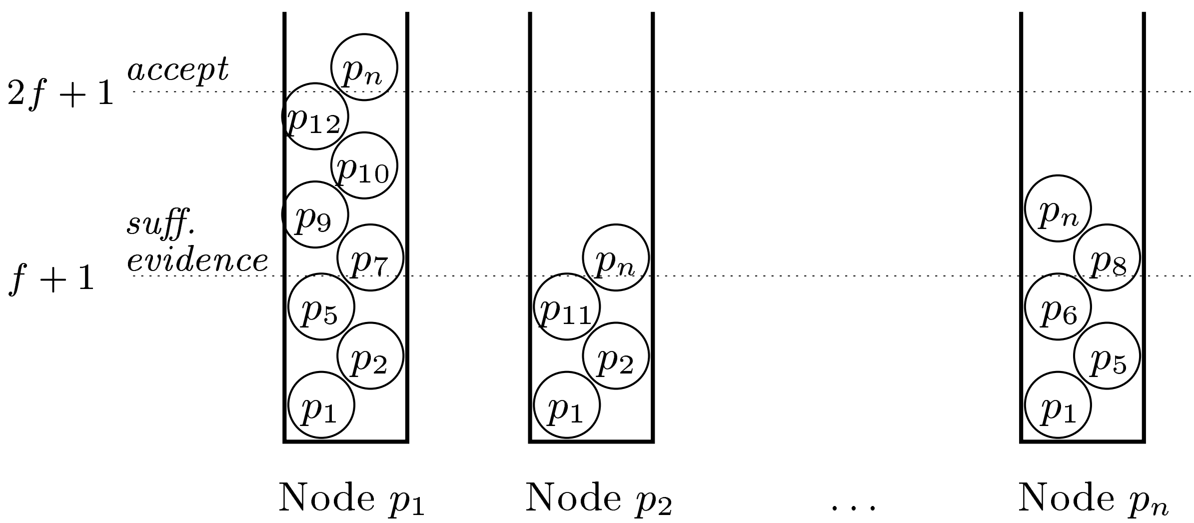
- Precision analysis depends only upon P1–P3

Consistent Broadcasting (Original Algorithm)

To broadcast msg, *each correct node:*
send *init_msg* to all;

To deliver msg, *each correct node:*
if received *init_msg* from $f+1$ distinct nodes
  $\rightarrow$ send *echo_msg* to all [once]; /* sufficient evidence */
$\square$ received *echo_msg* from $f+1$ distinct nodes
  $\rightarrow$ send *echo_msg* to all [once]; /* sufficient evidence */
fi
if received *echo_msg* from $2f+1$ distinct nodes
  $\rightarrow$ deliver *msg*; /* accept */
fi

Consistent broadcasting [Srikanth & Toueg, 1987]



Consistent Broadcasting (Improved Algorithm)

To broadcast msg*, each correct node:*
send *init_msg* to all;

To deliver msg*, each correct node:*
if received *init_msg* from $f_{\ell a} + f_a + f_s + f_{\ell s} + U + 1$ distinct nodes
→ send *echo_msg* to all [once]; /* sufficient evidence */
□ received *echo_msg* from $f_{\ell a} + f_a + f_s + f_{\ell s} + 1$ distinct nodes
→ send *echo_msg* to all [once]; /* sufficient evidence */
fi
if received *echo_msg* from $3f_{\ell a} + 2f_a + f_o + f_{\ell o} + f_s + f_{\ell s} + 1$ distinct nodes
→ deliver *msg*; /* accept */
fi

Improvement of Srikanth & Toueg's original consistent broadcasting primitive:

- Hybrid fault model for node faults (f_a, f_s, f_o, f_c)
- Hybrid fault model for link faults ($f_{\ell a}, f_{\ell s}, f_{\ell o}$)
- Unforgeability treshold U (triggering sensitivity)

Consistent Broadcasting (Events)

Event mapping:

- $I^s[t^s] \models$ "node s entered broadcast *msg*",
  $I^s_p[t^s_p] \models$ "node p received *init_msg* from s".
- $E^s[t^s] \models$ "node s achieved sufficient evidence",
  $E^s_p[t^s_p] \models$ "node p received *echo_msg* from s".

Admissible values and end-to-end delays:

- $I^s = I^s_p = \mathit{init\_msg}$
- $E^s = E^s_p = \mathit{echo\_msg}$
- $\boldsymbol{\tau}^s_p = [\tau^-, \tau^+]$ for all s, p, imposing
  - uncertainty $\varepsilon = \tau^+ - \tau^-$
  - maximum round-trip delay $t_{del} = 2\tau^+$
  - simple condition for symmetric faults

$$(t^s_q - \boldsymbol{\tau}^s_q) \cap (t^s_p - \boldsymbol{\tau}^s_p) \neq \emptyset \quad \Longleftrightarrow \quad |t^s_q - t^s_p| \leq \varepsilon$$

Consistent Broadcasting (Hybrid Fault Model)

Among each of the n pairs of perceptions

$$\left\{\begin{array}{c} I_p^s[t_p^s] \\ I_q^s \end{array}\right\} \in \left\{\begin{array}{c} \mathcal{I}_p(\infty) \\ \mathcal{I}_q(\infty) \end{array}\right\} \quad \text{resp.} \quad \left\{\begin{array}{c} E_p^s[t_p^s] \\ E_q^s \end{array}\right\} \in \left\{\begin{array}{c} \mathcal{E}_p(\infty) \\ \mathcal{E}_q(\infty) \end{array}\right\}$$

of any two non-faulty nodes p, q there may be at most

- f_c crash faults
- $f_s + 2f_{\ell s}$ symmetric faults, $2f_{\ell s}$ caused by at most $f_{\ell s}$ link faults ascribed to p resp. q each
- $f_o + 2f_{\ell o}$ omission faults, $2f_{\ell o}$ caused by at most $f_{\ell o}$ link faults ascribed to p resp. q each
- $f_a + 2f_{\ell a}$ asymmetric faults, $2f_{\ell a}$ caused by at most $f_{\ell a}$ link faults ascribed to p resp. q each

where the numbers are such that

$$n \geq 4f_{\ell a} + 3f_a + 2(f_o + f_{\ell o} + f_s + f_{\ell s}) + f_c + 1$$

Consistent Broadcasting (Results)

Theorem 1 (Properties Consistent Broadcast) *Under the perception-based hybrid fault model, the improved consistent broadcast primitive satisfies properties P1–P3 with*

- *delivery delay* $t_{del} = 2\tau^+$
- *acceptance delay* $t_\Delta = \tau^+ + \varepsilon = \tau^- + 2\varepsilon$

for any $0 \le U \le f_o + f_{\ell o}$ *and* $L = n - f_a - 2f_{\ell a} - f_o - f_{\ell o} + U$.

Comments:

- Our expression for t_Δ is more accurate than the one of [ST87]
- t_Δ could be further reduced (close to $\tau^- + \varepsilon$) by employing hardware timestamping **[DSN'01 Student paper by Martin Horauer]**
- For $U = 0$, *init_msg* and *echo_msg* could be merged in a single message, thereby reducing worst-case total number of messages from $2n^2$ to n^2

Conclusions

Accomplishments:

- Perception-based fault model covering node and link faults
- Hybrid classification of both time and value faults
- Application to consistent broadcasting revealed

  $$n \geq 4f_{\ell a} + 3f_a + 2(f_s + f_{\ell s} + f_o + f_{\ell o}) + f_c + 1$$

  sufficient for at most

  - f_a, f_s, f_o, f_c node faults
  - $f_{\ell a}$, $f_{\ell s}$, $f_{\ell o}$ link faults at any node

Consequences:

- Link faults easier to tolerate than node faults
- Large number $\Omega(n^2)$ of link faults tolerated systemwide (wireless applications!)

⇒ Transition from global to perception-based fault model worthwhile (and easy)

- Important further work: **Distributed consensus with link faults** [Schmid & Weiss, 2001] etc.