Computer and Automation Institute,
Hungarian Academy of Sciences

SOME MATHEMATICS FOR EVALUATION AND DESIGN OF COLLISION-RESOLVING STATEGIES IN NETWORKS

Third printing

Budapest 1985

Originally presented as an Institute Memo
under the same title in 1981
First reprint 1983

Készült

a Magyar Tudományos Akadémia Számitástechnikai
és Automatizálási Kutató Intézetében

Felelős kiadó:

DR VÁMOS TIBOR

ISBN 963 311 200 1

Hozott anyagról sokszorosítva
Utánnyomás

8616364 MTA Sokszorosító, Budapest. F. v.: dr. Héczey Lászlóné

CONTENTS

ABSTRACT

Ethernet-like strategies are investigated, to establish the time needed to resolve a collision (start a transmission) as a function of the number of competitors. Ethernet itself is shown to take an expected time proportional to the logarithm of that number. (Non-random strategies are worse.) Alternative strategies are proposed with bounded (and low) average.

0. PRELIMINARIES

"Ethernet-like" strategies will be considered. That is to mean, for our present purposes:

There is some *resource* (here, a cable or what). There is a number of *users* (stations), having equal access to the resource; they can *busy* it at any time but only a single one of them can *use* it (that is, successfully transmit) at a time. There is a time interval, called *slot*, , enough to enable any user to get informed of an action by another user; a user can detect if the resource is busy - that is, has been busied during the last slot. Moreover, a user can detect a *collision*, meaning two or more users busying the resource within a slot's time, or can at least detect its own colliding with others. - Every user has its own time, that is, slot-ticking; slots can be "synchronized" by events, that is, (beginnings or endings of) actions concerning the resource (which is in fact not really a synchronization, as different users may have detected the same event at different times - within a slot's time).

The following conventions are adhered to. While someone *is* transmitting, the others desist ("defer" their possible claims). Transmission termination (in practice, with some "recovery" time added) is a "synchronizer" for slots; stations take or leave actions in accordance with their slot clocking. As long as nothing happened, every station is free to start a transmission, and if it happened to be the only one to do so during a slot, it can proceed and transmit, the others having noticed, and started desisting. If not - in case of a collision - competitors choose, randomly, each for itself, a future slot to retry, and retry. If someone gets through (a station choosing the earliest slot is the only one that has chosen that slot), we are ready; else, repeat with a new choice. The process is evidently self-sustaining provided the retrial strategy is eventually able to resolve the collisions; the sooner, the better. (We speak of the system as existing from time immemorial, whereas it must be started, maybe new stations added, etc. Formally, at least one transmission must have been heard by anyone to get physically in touch with the system; this is to be waited for by a newcomer, the initial start having been made by a "creational" first transmission from somewhere. Not even this is really needed for the practical strategies we are going to examine; they will be seen to accommodate recruits automatically, with a trifling transient loss in efficiency.)

We consider systems where all participants follow the same strategy. (Except for transients, e.g., joining the system or station breakdown. We have to make sure that a strategy is "stable" in the sense that possible violations do not result in a collapse - nor in an enormous or protracted deterioration - of system functioning.) - A strategy will be regarded as a rule to decide upon a (re)trial slot somewhere in the future, which may depend on available system information (available at that particular station at that particular time). A randomized choice will be described as a decision upon the number of subsequent slots, any of which is to be selected with equal probability. (As a further strategical decision, the choice of a probability distribution might be admitted. But it can be

seen a priori that non-equiprobability is more collision-prone. Maybe, however, this "a priori" will become clearer in hindsight.)

To have an example of strategy, take "orthodox" Ethernet. Here every station starts in the very first slot it wants to *and* is permitted to. In case of a collision, the choice is made between the two following slots; if again collided, between the four following slots, and so on, over powers of two, between 2^k slots after the k^{th} collision. - Implementations used to restrict the number of doublings (this makes no essential difference). Sometimes the number of retrials is also restricted (in this case, "failure" is a possible outcome).

It is immediately clear (and will be shown) that this strategy works; and, in addition, it *has been* working for many years in many places.

Its underlying idea - combining slots and randomizing - is an amazingly bright concept, and perhaps the most thoroughgoingly "hierarchy-less" outfit now in being. It is qualitatively well-understood, and empirically well-examined. But it seems to be lacking, and in need of, a thorough mathematical scrutiny. Different strategies are possible, so some exact assessment of their efficiency would be desirable, both in themselves and in comparison (with each other and with quite different methods).

The point is, of course, to find the expected value of the number of slots used up in resolving the collisions.

In the first section, we deduce some formulae connecting this quantity with the number of *claimants* (stations wishing for transmission at a given moment) and the number of *lot-slots* (slots between which each claimant makes its own selection at each trial). In the second section, we cash the results.

1. FORMULAE

Let us denote the number of claimants by n, the number of lot-slots by m; $n,m \geq 1$.

PROBABILITY OF A SUCCESSFUL (COLLISION-FREE) TRIAL

An exhaustive and exclusive enumeration of success cases is the following.

- The first slot has been selected by exactly one claimant: 1XX...XXX. This can happen in $n(m-1)^{n-1}$ ways. (We have n choices to fill the first place, and n-1 remaining things to spread them in any way over the remaining m-1 places.)

- The first slot has been selected by none, the second by exactly one: 01X...XXX. This is $n(m-2)^{n-1}$ possibilities.

⋮

- The first m-2 slots have been selected by none, the second-to-last slot by exactly one: 000...01X. This is $n(m-(m-1))^{n-1} = n$ cases.

- Zero everywhere except one in the last slot: 000...001. This is obviously possible only in case n=1, and then in a single way. Stipulating, as usual, $0^0=1$, the expression $n(m-k)^{n-1}$ with $k=m$ comprises this case.

The number of collision-free cases is $n\sum_{k=1}^{m}(m-k)^{n-1} = n\sum_{k=0}^{m-1}k^{n-1}$; the number of all cases is m^n.

Denote $p_{nm} = \frac{n}{m^n}\sum_{k=0}^{m-1}k^{n-1}$. p_{nm} is the probability in question.

The probability of a collision is of course $1-p_{nm}$; denote it by q_{nm}.

Obviously, p_{nm} is non-decreasing for n fixed, m growing, and goes to 1. For m fixed, n growing, it is non-increasing and goes to 0. Especially, p_{n1} *is* 0 for $n \geq 2$.

THE EXPECTED SLOT OF SUCCESS OR COLLISION

The next questions are: *if* the trial results in a success, *which* (the how manieth) slot brings the success; *if* it results in a collision, *which* slot brings the collision. Let us denote these quantities by b_{nm} and c_{nm}, resp. – With $n=1$, c_{1m} is obviously undefined, and so is b_{n1} with $n \geq 2$.

For b_{nm}:

The success slot is the k^{th} slot if slots $1,\ldots, k-1$ have been selected by none, and slot k by a single claimant; the probability that the success occurs in that slot is $\frac{n(m-k)^{n-1}}{m^n p_{nm}}$ (the number of cases where that event occurs, divided by the number of all success cases). So

$$b_{nm} = \sum_{k=1}^{m} k\frac{n(m-k)^{n-1}}{m^n p_{nm}}$$

$$= \frac{n}{m^n p_{nm}}\left(m\sum_{k=1}^{m}(m-k)^{n-1} - \sum_{k=1}^{m}(m-k)^n\right)$$

$$= \frac{n}{m^n p_{nm}}\left(m\sum_{k=0}^{m-1}k^{n-1} - \sum_{k=0}^{m-1}k^n\right)$$

$$= \frac{n}{m^n p_{nm}}\left(\frac{m^{n+1}}{n}p_{nm} - \frac{m^{n+1}}{n+1}p_{n+1,m}\right)$$

$$= \frac{m\left(p_{nm} - \frac{n}{n+1}p_{n+1,m}\right)}{p_{nm}}.$$

For c_{nm}:

The collision slot is the k^{th} slot if slots $1,\ldots,k$ have been selected by none, and slot k by at least two claimants (that is, neither none nor one); the number of such cases is $(m-k+1)^n-(m-k)^n-n(m-k)^{n-1}$ (all go to the last $m-k+1$ slots, subtract the cases where all go to the last $m-k$ slots, and the cases where a single one goes to the k^{th} slot); the number of all collision cases is $m^n q_{nm} = m^n(1-p_{nm})$. So

$$c_{nm} = \sum_{k=1}^{m} k\frac{(m-k-1)^n-(m-k)^n-n(m-k)^{n-1}}{m^n q_{nm}}$$

$$= \frac{1}{m^n q_{nm}} \Big(m\Big(\sum_{k=1}^{m} (m-k-1)^n - \sum_{k=1}^{m} (m-k)^n - n \sum_{k=1}^{m} (m-k)^{n-1}\Big)$$

$$- \Big(\sum_{k=1}^{m} (m-k)(m-k-1)^n - \sum_{k=1}^{m} (m-k)^{n+1} - n \sum_{k=1}^{m} (m-k)^n\Big)\Big)$$

$$= \frac{1}{m^n q_{nm}} \Big(m\Big(\sum_{k=1}^{m} k^n - \sum_{k=0}^{m-1} k^n - n \sum_{k=0}^{m-1} k^{n-1}\Big)$$

$$- \Big(\sum_{k=1}^{m} (k-1)k^n - \sum_{k=0}^{m-1} k^{n+1} - n \sum_{k=0}^{m-1} k^n\Big)\Big)$$

$$= \frac{1}{m^n q_{nm}} \Big(m\Big(m^n - n \sum_{k=0}^{m-1} k^{n-1}\Big) - \Big(m^{n+1} - \sum_{k=1}^{m} k^n - n \sum_{k=0}^{m-1} k^n\Big)\Big)$$

$$= \frac{-mn \sum_{k=0}^{m-1} k^{n-1} + m^n + (n+1) \sum_{k=0}^{m-1} k^n}{m^n q_{nm}}$$

$$= \frac{-m^{n+1} p_{nm} + m^n + m^{n+1} p_{n+1,m}}{m^n q_{nm}}$$

$$= \frac{m(p_{n+1,m} - p_{nm}) + 1}{q_{nm}} .$$

– * –

We will need the expected number of the slot where anything happens, success or collision as the case is, the "decision slot". Let us denote it by a_{nm}.

Obviously

$$a_{nm} = p_{nm} b_{nm} + q_{nm} c_{nm}$$

$$= \frac{m}{n+1} p_{n+1,m} + 1$$

This could have been of course evaluated directly: the decision occurs at the k^{th} slot if everyone goes to the last $m-k+1$ slots, and at least one goes to slot k (that is, not everyone goes to the last $m-k$ slots); the number of all cases is m^n. So

$$a_{nm} = \frac{1}{m^n} \sum_{k=1}^{m} k((m-k+1)^n - (m-k)^n)$$

$$= \frac{1}{m^n} \sum_{k=1}^{m} (m-k+1)^n$$

$$= \frac{1}{m^n} \sum_{k=1}^{m} k^n$$

$$= \frac{m}{n+1} \Big(1+\frac{1}{m}\Big)^{n+1} p_{n+1,m+1} .$$

THE EXPECTED TRANSMISSION SLOT

The most general case is this. n_1 claimants scatter over m_1 slots, and try. If someone succeeds, transmission starts; else n_2 claimants scatter over m_2 slots, and retry - and so on, n_k claimants over m_k slots at the k^{th} trial, until transmission becomes possible, or in infinitum.

An immediate success occurs with probability $p_{n_1 m_1}$ and takes expectedly $b_{n_1 m_1}$ slots. Otherwise, with probability $q_{n_1 m_1} = 1 - p_{n_1 m_1}$, expectedly $c_{n_1 m_1}$ slots have been used up, and a retry follows, succeeding with a probability $p_{n_2 m_2}$; that is, a second-trial success has the overall probability $q_{n_1 m_1} p_{n_2 m_2}$, and takes expectedly $b_{n_2 m_2}$ additional slots. Else, with $q_{n_1 m_1} q_{n_2 m_2}$, additional $c_{n_2 m_2}$ is used up, and so on.

So we have

$$
\begin{aligned}
& p_{n_1 m_1} b_{n_1 m_1} \\
& + q_{n_1 m_1} c_{n_1 m_1} + q_{n_1 m_1} p_{n_2 m_2} b_{n_2 m_2} \\
& + \qquad q_{n_1 m_1} q_{n_2 m_2} c_{n_2 m_2} + q_{n_1 m_1} q_{n_2 m_2} p_{n_3 m_3} b_{n_3 m_3} \\
& + \ldots \\
& = \sum_{k=1}^{\infty} \prod_{h=1}^{k-1} q_{n_h m_h} \left(p_{n_k m_k} b_{n_k m_k} + q_{n_k m_k} c_{n_k m_k} \right),
\end{aligned}
$$

an empty product meaning 1.

The expression in parentheses is $a_{n_k m_k}$, so the expected transmission slot can be written

$$\sum_{k=1}^{\infty} \prod_{h=1}^{k-1} q_{n_h m_h} a_{n_k m_k}.$$

Certain simplifications offer themselves immediately. n_k being the number of claimants, will not undergo violent changes during the contest for transmission. It could be forced to remain constant during the contest, simply by imposing desistment (not only for transmission but) for contest periods, which is physically possible; thus those and only those would participate in a contest who started it (except if someone breaks down in the meantime). But even if we allow for newcomers in a contest, there is a low probability of someone's dropping in amidst a contest, since contest periods take but a (hopefully) small proportion of the overall time. And since no transmission takes place during contest, claims will not vanish (if nobody goes down). So changes in n_k during a contest, even if permitted, may be safely neglected, and the subscript of n may be dropped, which gives

$$\sum_{k=1}^{\infty} \prod_{h=1}^{k-1} q_{n m_h} a_{n m_k}.$$

The case of m_k is more delicate: It may be (and is) a meaningful strategy to change m_k from trial to trial. (So does Ethernet.)

Nevertheless, the case of an unvaried m will prove to be of paramount importance, and we are going to write it down, denoting it by d_{nm}.

$$d_{nm} = \sum_{k=1}^{\infty} q_{nm}^{k-1} a_{nm}.$$

For $n \geq 2$, $m=1$ - we call this a "hopeless trial" - d_{nm} is infinite (transmission never starts).

In all other cases $p_{nm} > 0$, so $q_{nm} < 1$; $\sum_{k=1}^{\infty} q_{nm}^{k-1} = \frac{1}{1-q_{nm}} = \frac{1}{p_{nm}}$, and we have

$$d_{nm} = \frac{a_{nm}}{p_{nm}} = \frac{\frac{m}{n+1} p_{n+1,m+1}}{p_{nm}} = \frac{m}{n+1}\left(1+\frac{1}{m}\right)^{n+1} \frac{p_{n+1,m+1}}{p_{nm}}.$$

d_{nm} has a minimum as a function of $m=1,2,\ldots$. To see this, it is enough to see that it cannot take an infinite sequence of decreasing values as $m \to \infty$, which is clear from its rightmost expression, since $p_{nm} \to 1$ for $m \to \infty$.

Coming back to the more general formula: We call an "admissible choice-sequence" any sequence m_k for which transmission will start with probability 1. If an admissible choice sequence contains hopeless trial terms (terms with $m_k=1$ for $n \geq 2$) then the sequence with those terms left out is also admissible and the value of the formula is not greater than for the original

sequence. For such a sequence, we can write - using $a_{nm_k} = p_{nm_k} d_{nm_k}$ -:

$$\sum_{k=1}^{\infty} \prod_{h=1}^{k-1} q_{nm_h} p_{nm_k} d_{nm_k}.$$

For all admissible choice-sequences we have

$$\sum_{k=1}^{\infty} \prod_{h=1}^{k-1} q_{nm_h} p_{nm_k} = 1,$$

since this is the sum of the probabilities that transmission starts at the first, second, ..., k^{th}, ... trial.

Let d denote an arbitrary constant, and write $d_{nm_k} = d + e_{nm_k}$. Then the formula turns to

$$d + \sum_{k=1}^{\infty} \prod_{h=1}^{k-1} q_{nm_h} p_{nm_k} e_{nm_k}.$$

Now set d to the minimal value of d_{nm}. Then all the e_{nm_k} are non-negative, and so are of course their coefficients. It turns out that no choice-sequence can perform better than taking for m_k invariably in each trial the particular value of m for which d_{nm} is minimal.

(Note that we speak here of choice-sequences, not of strategies. A strategy produces a choice-sequence, so it cannot perform better than that. But a choice-sequence has not necessarily a strategy that produces it, so it is not clear from the outset whether there exists a strategy that reaches or approaches that limit.)

WHAT MUCH IS THAT MUCH?

Our formulae are very nice, but not very handy. We are going to deduce more tractable expressions for them.

The "Bernoulli numbers" are the numbers B_k defined by the equation

$$\frac{t}{e^t-1} = \sum_{k=0}^{\infty} \frac{B_k}{k!}t^k .$$

This *is* a definition, since the right-hand side is a Taylor series, and a Taylor series determines its coefficients uniquely.

Note that $\frac{t}{e^t-1}$ is regular in $t=0$, and its nearest singularities are at $t=\pm 2\pi i$, so the series on the right converges to $\frac{t}{e^t-1}$ in the circle $|t|<2\pi$.

First we show that

$$n\sum_{k=0}^{m-1} k^{n-1} = \sum_{k=0}^{n-1} \binom{n}{k} B_k m^{n-k} .$$

We express $\frac{t(e^{mt}-1)}{e^t-1}$ in two ways.

On one hand,

$$\frac{t(e^{mt}-1)}{e^t-1} = \frac{t}{e^t-1}e^{mt} - \frac{t}{e^t-1}$$

$$= \sum_{n=0}^{\infty} \frac{B_n}{n!}t^n \sum_{n=0}^{\infty} \frac{m^n t^n}{n!} - \sum_{n=0}^{\infty} \frac{B_n}{n!}t^n$$

$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{B_k}{k!} \frac{m^{n-k}}{(n-k)!}t^n - \sum_{n=0}^{\infty} B_n \frac{t^n}{n!}$$

$$= \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} \binom{n}{k} B_k m^{n-k} - B_n \right) \frac{t^n}{n!}$$

$$= \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n-1} \binom{n}{k} B_k m^{n-k} \right) \frac{t^n}{n!} .$$

On the other hand,

$$\frac{t(e^{mt}-1)}{e^t-1} = t\frac{e^{mt}-1}{e^t-1}$$

$$= t\sum_{k=0}^{m-1} e^{kt} = t\sum_{k=0}^{m-1} \sum_{n=0}^{\infty} \frac{k^n t^n}{n!}$$

$$= t\sum_{n=0}^{\infty} \sum_{k=0}^{m-1} k^n \frac{t^n}{n!}$$

$$= \sum_{n=0}^{\infty}\left((n+1)\sum_{k=0}^{m-1} k^n\right)\frac{t^{n+1}}{(n+1)!}$$

$$= \sum_{n=1}^{\infty}\left(n\sum_{k=0}^{m-1} k^{n-1}\right)\frac{t^n}{n!}\ .$$

(Changes in summation order are justified by absolute convergence.)

Comparing co-efficients of $\frac{t^n}{n!}$ gives the result.

Consequently,

$$p_{nm} = \frac{n}{m^n}\sum_{k=0}^{m-1} k^{n-1}$$

$$= \sum_{k=0}^{n-1}\binom{n}{k}\frac{B_k}{m^k}\ .$$

Note that $m^{n-1}p_{nm}$ is an ordinary polynomial in m. Especially, the formula is defined for non-integer values of m. We will use this fact in getting estimates, but the values we are interested in are those for positive integral arguments. (Non-integral arguments have no probabilistic interpretation, not even in a sense of "mixing": e.g., p_{5m} is negative in the immediate right neighbourhood of m=1.)

For small values of n, p_{nm} can be explicitly written down by recursively computing the B_k.

We have B_0=1 (evidently), and $B_k = -\frac{1}{k+1}\sum_{h=0}^{k-1}\binom{k+1}{h}B_h$ for $k\geq 1$ (to see this, rewrite the relation in the form $\sum_{h=0}^{k}\binom{k+1}{h}B_h=0$, and set m=1 in the first expression for $t\frac{e^{mt}-1}{e^t-1}$). Hence $B_1=-\frac{1}{2}$, $B_2=\frac{1}{6}$, $B_3=0$, $B_4=-\frac{1}{30}$, etc.

$$p_{1m} = 1,$$

$$p_{2m} = 1 - \frac{1}{m},$$

$$p_{3m} = 1 - \frac{3}{2m} + \frac{1}{2m^2} = (1-\frac{1}{m})(1-\frac{1}{2m}),$$

$$p_{4m} = 1 - \frac{2}{m} + \frac{1}{m^2} = (1-\frac{1}{m})^2,$$

$$p_{5m} = 1 - \frac{5}{2m} + \frac{5}{3m^2} - \frac{1}{6m^4}$$

$$= (1-\frac{1}{m})(1-\frac{1}{2m})(1-\frac{1}{m}-\frac{1}{3m^2}),$$

etc.

We are going to investigate p_{nm} for n growing. (This is not to pretend that n may grow towards infinity; n is limited by the number of stations on the network, which is itself not a very large number after all. But the results will be relevant for small values of n, too.) We will be interested in cases where m approximately equals n, and we assume $n<2\pi m$ from the outset, and even $\frac{n}{m}\leq 2\pi-\varepsilon$ with some fixed positive ε.

$$p_{nm} = \sum_{k=0}^{n-1} \binom{n}{k} \frac{B_k}{m^k} = \sum_{k=0}^{n} \binom{n}{k} \frac{B_k}{m^k} - \frac{B_n}{m^n};$$

the upper sum-limit can be set ∞ instead of n; so

$$p_{nm} = \sum_{k=0}^{\infty} \frac{B_k}{k!} \left(\frac{n}{m}\right)^k + r_{nm}$$

where

$$r_{nm} = -\sum_{k=0}^{\infty} \left(\frac{n^k}{k!} - \binom{n}{k}\right) \frac{B_k}{m^k} - \frac{B_n}{m^n}.$$

Looking at the defining formula of the B_k, we find that

$$p_{nm} = \frac{\frac{n}{m}}{e^{\frac{n}{m}} - 1} + r_{nm}.$$

Note that the first term depends on n and m only via $\frac{n}{m}$.

For r_{nm}, the lower sum-limit can be set 2 instead of 0, and we have

$$|r_{nm}| \le \sum_{k=2}^{\infty} \left|\frac{n^k}{k!} - \binom{n}{k}\right| \frac{|B_k|}{m^k} + \frac{|B_n|}{m^n}.$$

Using $\frac{n^k}{k!}\left(1 - \frac{k(k-1)}{2n}\right) \le \binom{n}{k} \le \frac{n^k}{k!}$ (the upper bound is trivial, the lower bound is easily seen to hold by induction on k), this goes to

$$|r_{nm}| \le \frac{1}{2n} \sum_{k=2}^{\infty} k(k-1) \frac{|B_k|}{k!} \left(\frac{n}{m}\right)^k + \frac{|B_n|}{m^n}.$$

The radius of convergence of $\sum_{k=0}^{\infty} \frac{B_k}{k!} t^k$ being 2π, we have $\limsup_{k\to\infty} \sqrt[k]{\frac{|B_k|}{k!}} = \frac{1}{2\pi}$. Therefore $\limsup_{k\to\infty} \sqrt[k]{k(k-1)\frac{|B_k|}{k!}} = \frac{1}{2\pi}$ too, since $\lim_{k\to\infty} \sqrt[k]{k(k-1)} = 1$. Thus the sum converges for $\frac{n}{m} < 2\pi$, and converges uniformly for $\frac{n}{m} \le 2\pi - \varepsilon$, it is therefore continuous and has there a finite maximum. So $|r_{nm}| \le \frac{\text{const.}}{n} + \frac{|B_n|}{m^n}$. - For $\frac{|B_n|}{m^n}$, we may rewrite the limsup relation as $\frac{|B_n|}{n!} \le \left(\frac{1}{2\pi} + \varepsilon_n\right)^n$ with $\varepsilon_n \to 0$ for $n \to \infty$. $\frac{|B_n|}{m^n} = \frac{|B_n|}{n!} \frac{n!}{m^n}$; using $m \ge \frac{n}{2\pi - \varepsilon}$ and the Stirling formula, we get $\frac{|B_n|}{m^n} \le \text{const.} \frac{\sqrt{n}}{e^n} \le \frac{\text{const.}}{n}$ (with a second and third "const.").

- All in all,

$$|r_{nm}| \le \frac{\text{const.}}{n} \quad \text{if} \quad \frac{n}{m} \le 2\pi - \varepsilon.$$

In addition, we will need an upper bound on p_{nm} for all n, m, especially for $n \gg m$. We obtain this from the first formula for p_{nm}. p_{1m} is 1. For $n > 1$,

$$p_{nm} = \frac{n}{m^n} \sum_{k=1}^{m-1} k^{n-1}$$

$$= \frac{n}{m}\left(\left(1-\frac{1}{m}\right)^{n-1} + \left(1-\frac{2}{m}\right)^{n-1} + \dots + \left(1-\frac{m-1}{m}\right)^{n-1}\right).$$

Write

$v = \frac{n-1}{m}$, then this is

$$= \frac{n}{n-1}\, v\left(\left(1-\frac{v}{n-1}\right)^{n-1} + \left(1-\frac{2v}{n-1}\right)^{n-1} + \dots + \left(1-\frac{\left(\frac{n-1}{v}-1\right)v}{n-1}\right)^{n-1}\right).$$

All terms in the last factor have the form $(1-\frac{x}{h})^h$ with $\frac{x}{h}<1$; so $1-\frac{x}{h}<e^{-\frac{x}{h}}$ and $(1-\frac{x}{h})^h<e^{-x}$, which yields the estimate

$$\leq \frac{n}{n-1} v\left(e^{-v}+e^{-2v}+\dots e^{-\left(\frac{n-1}{v}-1\right)v}\right)$$

$$= \frac{n}{n-1} v e^{-v} \frac{1-e^{-(n-1-v)}}{1-e^{-v}}$$

$$\leq \frac{n}{n-1} \frac{ve^{-v}}{1-e^{-v}}$$

$$= \frac{n}{n-1} \frac{v}{e^{v}-1}.$$

Putting in the value of v,

$$p_{nm} \leq \frac{\frac{n}{m}}{e^{\frac{n-1}{m}}-1}$$

Now we go to the d_{nm}.

First for small n. By explicit computation,

$$d_{1m} = \frac{m+1}{2},$$

$$d_{2m} = \frac{(2m+1)(m+1)}{6(m-1)},$$

$$d_{3m} = \frac{m(m+1)^2}{2(m-1)(2m-1)},$$

$$d_{4m} = \frac{(m+1)(2m+1)(3m^2+3m-1)}{30(m-1)^2 m},$$

etc.

Minimum values:

d_{1m} is monotonous, its minimum value is assumed at the left endpoint of the domain: $m=1$, and is 1.

d_{2m} has its minimum at $m=1+\sqrt{3} \approx 2.73$; it is $\frac{7}{6}+\frac{2}{\sqrt{3}} \approx 2.3214$. For integral m, the minimum is at $m=3$, $d_{23}=\frac{7}{3}=2.\overline{3}$.

d_{3m}: minimum at ≈ 3.54, is ≈ 2.3624. Integral minimum: $d_{34}=\frac{50}{21} \approx 2.3810$.

d_{4m}: minimum at ≈ 4.54, is ≈ 2.4360. Integral minimum: $d_{45} = \frac{979}{400} = 2.4475$.

(For more values see the Appendix.)

In the general case,

$$d_{nm} = \frac{m}{n+1}\left(1+\frac{1}{m}\right)^{n+1}\frac{p_{n+1,m+1}}{p_{nm}}.$$

For $\frac{n}{m} \leq 2\pi-\varepsilon$, using $p_{nm} = \frac{\frac{n}{m}}{e^{\frac{n}{m}}-1} + r_{nm}$ with $|r_{nm}| \leq \frac{\text{const.}}{n}$, we see that $n\to\infty$, $\frac{n}{m}\to v$ implies that $\frac{n}{m}$, $\frac{n+1}{m}$, $\frac{n+1}{m+1}$ all go to v, $\frac{m}{n+1}$ goes to $\frac{1}{v}$, $(1+\frac{1}{m})^{n+1}$ goes to e^v, both p_{nm} and $p_{n+1,m+1}$ go to $\frac{v}{e^v-1}$, their quotient goes to 1, so d_{nm} goes to $\frac{e^v}{v}$, and all this uniformly. Thus, for n large enough, and m near enough $\frac{n}{v}$, d_{nm} comes as close to $\frac{m}{n}e^{\frac{n}{m}} \sim \frac{e^v}{v}$ as wished. - For $v>0$, $\frac{e^v}{v}$ has its minimum at $v=1$ (differentiate it to find the local minimum, and make sure that there are no non-local minima), the minimum value is $e \approx 2.7183$. So, for n large, d_{nm} takes minimum values close to e, and this at locations m close to n.

So far, we considered only the range $\frac{n}{m} \leq 2\pi-\varepsilon$; maybe better minima could be found outside that range. This is not the

case. Replace p_{nm} by its upper estimate, and write out $p_{n+1,m+1}$ explicitly, then

$$d_{nm} \geq \frac{m}{n}\left(e^{\frac{n-1}{m}}-1\right)\frac{m^n+(m-1)^n+\ldots+1^n}{m^n}.$$

The last factor being ≥ 1,

$$d_{nm} \geq \frac{m}{n}\left(e^{\frac{n-1}{m}}-1\right).$$

Now take $m \leq \frac{n}{6}$ (which comprises the range $m < \frac{n}{2\pi-\varepsilon}$). Then we have $d_{nm} \geq \frac{1}{6}(e^3-1) > e$ for $n \geq 2$. (The case $n=1$ has been explicitely treated.)

So the minimum of d_{nm} starts with the values 1, 2.3333^+, 2.3810^-, 2.4475 for $n=1,2,3,4$, and approaches 2.7183^- for $n\to\infty$.

We have the (perhaps surprising) result that the m-minimum of d_{nm}, as a function of n, stays below some fixed bound, that is, an appropriate choice of the lot-slots can resolve collisions in limited expected time, whatever the number of conflicting claims is.

We are not yet assured that this minimum is not going to make great excursions before it settles for its limiting value, because, for all we know, the "const." in our estimate for $|r_{nm}|$ can be as large as anything. Intuitively, it looks fairly obvious, that the minimum is increasing for growing n, which would settle the problem. Nonetheless, we wish for a possibly sharp estimate, so much the more as non-optimal, merely "near-optimal" m-values, too, will be of interest.

First we show: For $k \geq 1$,

$$B_{2k} = (-1)^{k-1} \frac{2(2k)!}{(2\pi)^{2k}} \zeta(2k),$$

$$B_{2k+1} = 0,$$

where ζ is the Riemann Zeta function, $\zeta(2k) = \sum_{n=1}^{\infty} \frac{1}{n^{2k}}$.

We have

$$\frac{t}{e^t-1} + \frac{t}{2} = \frac{t}{2} \frac{e^t+1}{e^t-1} = \frac{t}{2} \frac{e^{\frac{t}{2}} + e^{-\frac{t}{2}}}{e^{\frac{t}{2}} - e^{-\frac{t}{2}}} = \frac{t}{2} \coth \frac{t}{2}.$$

$\coth t$ is the logarithmic derivative of $\sinh t$.

We write $\sinh t$ in its "product form"

$$\sinh t = t \prod_{n=1}^{\infty} \left(1 + \left(\frac{t}{2\pi n}\right)^2\right)$$

(which is immediately obtained from the analogous product form for $\sin t$, to be found in practically any handbook of Analysis).

Herefrom, for $|t| < 2\pi$,

$$\log \sinh t = \log t + \sum_{n=1}^{\infty} \log \left(1 + \left(\frac{t}{2\pi n}\right)^2\right),$$

$$\coth t = \frac{1}{t} + 2 \sum_{n=1}^{\infty} \frac{\frac{t}{(2\pi n)^2}}{1 + \left(\frac{t}{2\pi n}\right)^2} = \frac{1}{t} + 2 \sum_{n=1}^{\infty} \frac{t}{t^2 + (2\pi n)^2}.$$

So

$$\frac{t}{e^t-1} = \frac{t}{2} \coth \frac{t}{2} - \frac{t}{2} = 1 - \frac{t}{2} + 2 \sum_{n=1}^{\infty} \frac{t^2}{t^2 + (2\pi n)^2}.$$

Now

$$\frac{t^2}{t^2 + (2\pi n)^2} = \sum_{k=1}^{\infty} (-1)^{k-1} \left(\frac{t}{2\pi n}\right)^{2k}.$$

Putting this in, and interchanging summations (permitted by absolute convergence), we get

$$\frac{t}{e^t-1} = 1 - \frac{t}{2} + \sum_{k=1}^{\infty} (-1)^{k-1} \frac{2}{(2\pi)^{2k}} \sum_{n=1}^{\infty} \frac{1}{n^{2k}} t^{2k}.$$

Comparison of co-efficients with the defining series of the B_k gives the result.

The estimate for $|r_{nm}|$ can now be sharpened.

As, for $k \geq 1$, B_{2k+1} is 0 and B_{2k} alternates in sign, we see that the sum in the estimate is

$-t^2 \frac{d^2}{dt^2} \frac{t}{e^t-1}$ taken at $t = i\frac{n}{m}$.

$$-t^2 \frac{d^2}{dt^2} \frac{t}{e^t-1} = -t^2 \frac{(t-2)e^{2t} + (t+2)e^t}{(e^t-1)^3}$$

$$= -t^2 \frac{(t-2)e^{\frac{t}{2}} + (t+2)e^{-\frac{t}{2}}}{(e^{\frac{t}{2}} - e^{-\frac{t}{2}})^3}$$

$$= -t^2 \frac{t \cosh \frac{t}{2} - 2 \sinh \frac{t}{2}}{4 \sinh^3 \frac{t}{2}}.$$

Substituting $t = i\frac{n}{m}$, we have for the sum

$$\left(\frac{n}{m}\right)^2 \frac{2\sin\frac{n}{2m} - \frac{n}{m}\cos\frac{n}{2m}}{4\sin^3\frac{n}{2m}} .$$

$\frac{|B_n|}{m^n}$ is 0 for odd $n>1$, otherwise, by Stirling's formula, it is $\le \sqrt{8\pi}\,\zeta(n)\,\sqrt{n}\left(\frac{n}{2\pi e m}\right)^n e^{\frac{1}{12n}}$; regrouping this,

$$\frac{|B_n|}{m^n} \le \left(\frac{\sqrt{2}}{e\sqrt{\pi}}\,\zeta(n)\,e^{\frac{1}{12n}}\,n^{\frac{3}{2}}\left(\frac{n}{2\pi e m}\right)^{n-1}\right)\frac{1}{m} .$$

We will suppose $n>4$, $\frac{n}{m} \le \pi$.

$\zeta(n) = \sum_{k=1}^{\infty} \frac{1}{k^n}$ is evidently decreasing for growing n, and so is the whole parenthesized first factor. Within it, we can take $n=6$ (B_5 being 0). $\zeta(6) = \frac{\pi^6}{945}$. (Compute B_6 from the recursive formula - it is $\frac{1}{42}$ -, and use the formula connecting $\zeta(6)$ and B_6.) With this and $\frac{n}{m} \le \pi$, we get

$$\frac{|B_n|}{m^n} \le \frac{0.001}{m} .$$

Putting things together, we have, for $n>4$, $\frac{n}{m} \le \pi$,

$$|r_{nm}| \le \frac{1}{n}\,\frac{n}{m}\left(\frac{n}{m}\,\frac{2\sin\frac{n}{2m} - \frac{n}{m}\cos\frac{n}{2m}}{8\sin^3\frac{n}{2m}} + 0.001\right).$$

Now we estimate d_{nm} from above. We use

$$d_{nm} = \frac{\frac{m}{n+1}\,p_{n+1,m} + 1}{p_{nm}} .$$

We replace $p_{n+1,m}$ by the upper estimate and p_{nm} by $\frac{\frac{n}{m}}{e^{\frac{n}{m}}-1} - |r_{nm}|$. So

$$d_{nm} \le \frac{m}{n}\,e^{\frac{n}{m}}\left(1 - (e^{\frac{n}{m}} - 1)\,\frac{m}{n}\,|r_{nm}|\right)^{-1} .$$

We denote

$$s_{nm} = (e^{\frac{n}{m}} - 1)\left(\frac{n}{m}\,\frac{2\sin\frac{n}{2m} - \frac{n}{m}\cos\frac{n}{2m}}{8\sin^3\frac{n}{2m}} + 0.001\right).$$

With this,

$$d_{nm} \le \frac{m}{n}\,e^{\frac{n}{m}}\left(1 - \frac{s_{nm}}{n}\right)^{-1} .$$

s_{nm} clearly decreases if $\frac{n}{m}$ decreases. In particular,

$$s_{nm} \le 1.622 \text{ for } \tfrac{n}{m} \le 2,$$
$$s_{nm} \le 0.5517 \text{ for } \tfrac{n}{m} \le \tfrac{3}{2},$$
$$s_{nm} \le 0.1604 \text{ for } \tfrac{n}{m} \le 1.$$

Considering $n>4$, the factor $\left(1 - \frac{s_{nm}}{n}\right)^{-1}$ is

$$\le 1.4802 \text{ for } \tfrac{n}{m} \le 2,$$
$$\le 1.1241 \text{ for } \tfrac{n}{m} \le \tfrac{3}{2},$$
$$\le 1.0332 \text{ for } \tfrac{n}{m} \le 1.$$

Especially,

$$d_{nn} \le 1.0332\, e < 2.81 .$$

For $n \le 4$, the explicit forms show that the same relation (amply) holds for d_{nn}.

As $\min_m d_{nm} \le d_{nn}$, $\min_m d_{nm} < 2.81$ for all n. (More is true, but it is not worth the effort to show it.)

We search for "good intervals" for m.

For $m = \frac{4}{3} n$ we have $d_{nm} \sim \frac{4}{3} e^{\frac{3}{4}} < 2.823$; for $m = \frac{2}{3} n$ we have $d_{nm} \sim \frac{2}{3} e^{\frac{3}{2}} < 2.988$, and for values in between d_{nm} is not greater than at the worse endpoint. Considering the "tolerances", we have practically $d_{nm} \le 3$ with such m for $n > 4$, and the explicit forms show the same for $n \le 4$. (Note that the interval $[\frac{2}{3} n, \frac{4}{3} n]$ contains a power of two.)

Even for $m \in [\frac{2}{3} n, \frac{8}{3} n]$ d_{nm} remains well under 4.

The <u>best</u> value of m is 1 for $n=1$, and $n+1$ for $n \ge 2$ as long as computations go (until 25); maybe this holds through. Anyway, $\frac{n}{n+1}$ goes to 1, so $d_{n,n+1}$ goes to e, and we have the same (in fact a somewhat better) overall upper limit for it as for d_{nn}. (For small n, d_{nn} markedly exceeds $d_{n,n+1}$, but it too is very nice.)

(If we (that is, the particular stations) <u>knew</u> n, this would give a strategy (with the expectation of starting a transmission well within 3 slot times, irrespective of the number of competitors). But n is not known, so this is not a strategy. Yet, conversely, a well-chosen m_k implies a good guess on n. This suggests to regard strategies somehow as means for "guessing n", the better

the guessing the better the strategy. Imprecise as this remark is, it will have exact implications in the next section. But we have to deal with another point before.)

TRIAL OVERHEAD

Thus far, we tacitly assumed that trials take no extra time over and above the slots they use up, an assumption that might prove inadmissible. E.g., the cable may require some recovery time after each "action" (transmission or trial) which is not negligible as compared with slot time.

Let w denote that extra time per action, expressed in slot units. $w \ge 0$; in practice, $w \le 5$.

Its impact: b_{nm}, c_{nm} get simply incremented by w, and so does a_{nm}.

So the value of the "general formula" increases by

$$w \sum_{k=1}^{\infty} \prod_{h=1}^{k-1} q_{nm_h} .$$

We examine the effects on d_{nm}.

Denoting by $d^{(w)}_{nm}$ the expected transmission start time with trial overhead times included, we have obviously

$$d^{(w)}_{nm} = d_{nm} + \frac{w}{p_{nm}} .$$

For small n,

$$d_{1m}^{(w)} = \frac{m+1}{2} + w.$$ The minimum remains at $m=1$ and becomes $1+w$.

$$d_{2m}^{(w)} = \frac{m}{3} + \frac{5}{6} + \frac{1+mw}{m-1}$$. The minimum is at $m = 1+\sqrt{3(1+w)}$ and is $\frac{7}{6} + \frac{2}{\sqrt{3}}\sqrt{1+w} + w$. Integral minimum locations are at $m=3$ for $0\le w\le 1$, at $m=4$ for $1\le w\le 3$, at $m=5$ for $3\le w\le 5.\overline{6}$; the minima are $2.\overline{3}$ for $w=0$, $3.8\overline{3}$ for $w=1$, $5.1\overline{6}$ for $w=2$, 6.5 for $w=3$, 7.75 for $w=4$, 9 for $w=5$.

Etc.

For $n\to\infty$, $\frac{n}{m}\to v$,

$$d_{nm}^{(w)} \sim \frac{e^v}{v} + w\frac{e^v-1}{v} = \frac{e^v(1+w)-w}{v} .$$

Convergence speed can be estimated through $|r_{nm}|$ the same way as above.

The delay caused by trial overhead is at least w (since there is at least one trial), its minimum is not more in the limit than $(e-1)w<1.72w$ (since this comes out if we take $v=1$ as if nothing happened). It can be both felt and shown that a positive w shifts the overall optimum to a greater m (smaller v). As things are not very sensitive, we put up with kitchen methods.

By differentiation, we see that what we need for a minimum is

$$(v-1)\, e^v(1+w) + w = 0.$$

Solving this for w:

$$w = \frac{(1-v)e^v}{1-(1-v)e^v} .$$

(It *could* be solved for v, e.g. by

$$v = 1 - \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} \left(\frac{1}{e} \frac{w}{1+w}\right)^k ,$$

which we leave unproved and unused.)

Now a rule of thumb is this. If $w<\frac{2\sqrt{e}-e}{(\sqrt{e}-1)^2} \approx 1.376$ then $v=1$ is better than $v=\frac{1}{2}$; otherwise $v=\frac{1}{2}$ is better than $v=1$; anything in between is tolerable.

Indeed, $v=1$ is optimal for $w=0$ (and yields $e \approx 2.718$): $v=\frac{1}{2}$ is optimal for $w=\frac{\sqrt{e}}{2-\sqrt{e}} \approx 4.693$ (and yields $\frac{2\sqrt{e}}{2-\sqrt{e}} \approx 9.387 \approx e+1.421w \approx 2\sqrt{e}+1.297w$); for w on the watershed, $v=1$ and $v=\frac{1}{2}$ yield the same value.

(Some numerical data: $w=1$ with $v=1$ gives $2e-1<4.437$; $w=5$ with $v=\frac{1}{2}$ gives $12\sqrt{e}-10<9.785$. This is 2.6% worse than the optimum for $w=1$, and even better for $w=5$.

Even in the middle, the loss is well under 5%. - Optima have been estimated by numerical computation, not reproduced here. - If w happens to be >5, $v=\frac{1}{2}$ continues of course to yield $2\sqrt{e}+2(\sqrt{e}-1)w \approx 3.297+1.297w$, but this ceases to be optimal: for a high w, a somewhat better v can be found along the same lines.)

2. STRATEGIES

First we investigate "orthodox" Ethernet, then we consider other possibilities.

In what follows, the lot-slot number m will be mostly a power of two. We adopt the convention that an upper-case letter denotes 2 raised to the value of the corresponding lower-case letter: e.g., $H = 2^h$. - lg will denote the logarithm to the base 2.

ETHERNET

Untruncated Ethernet strategy yields in case of n claimants the expected transmission slot

$$e_n = \sum_{k=0}^{\infty} \prod_{h=0}^{k-1} q_{nH} a_{nK},$$

since it starts with a single lot-slot and doubles the number at every collision.

Doubling limitation means that there is some u so that min (H,U) and min (K,U) come into the subscripts instead of H and K, resp. - Retrial limitation means that the sum goes to some u instead of ∞; e_n is then interpreted as the expected slot in the non-failure case (having the probability $\sum_{k=0}^{u} \prod_{h=0}^{k-1} q_{nH} p_{nK}$).

We start with no limitations.

Values of e_n with small n are easily computed:

$$e_1 = 1,$$
$$e_2 = 3.6888^+,$$
$$e_3 = 3.8527^+,$$
$$e_4 = 4.1609^+,$$

etc.

For large n, we begin with estimating e_n from below.

Since the terms of the sum are non-negative, we can omit its tail from some u on (to be fixed later):

$$e_n \geq \sum_{k=0}^{u} \prod_{h=0}^{k-1} q_{nH} a_{nK}.$$

Since $a_{nm} \geq 1$ (as $a_{nm} = \frac{m}{n+1} p_{n+1,m} + 1$, and $p_{n+1,m} \geq 0$),

we can drop the a_{nK}:

$$e_n \geq \sum_{k=0}^{u} \prod_{h=0}^{k-1} q_{nH}.$$

Since p_{nm} increases for m growing, $q_{nm} = 1 - p_{nm}$ decreases; so we can replace the q_{nH} by q_{nU}:

$$e_n \geq \sum_{k=0}^{u} q_{nU}^{k}.$$

Let $\hat{p}_{nm}$ denote $\dfrac{\frac{n}{m}}{\exp\frac{n-1}{m} - 1}$. We know that $p_{nm} \leq \hat{p}_{nm}$; so $q_{nU} = 1 - p_{nU}$ can be replaced by $1 - \hat{p}_{nU}$:

$$e_n \geq \sum_{k=0}^{u} (1 - \hat{p}_{nU})^k$$

$$= \frac{1 - (1 - \hat{p}_{nU})^{u+1}}{\hat{p}_{nU}}$$

$$\leq \frac{1 - \exp(-(u+1)\hat{p}_{nU})}{\hat{p}_{nU}}.$$

(We used that $\hat{p}_{nU} \neq 0$ and that $1 - x \leq e^{-x}$.)

Now we proceed in two steps.

First, set

$$u = \left\lfloor \lg \frac{n-1}{\log(\lambda \lg n) + \log\log(\lambda \lg n)} \right\rfloor$$

with some λ (not yet specified but large enough so that the loglog make sense).

Substituting u (do it), we get for n growing:

$$\hat{p}_{nU} \sim \frac{1}{\lambda \lg n},$$

$$(u+1)\hat{p}_{nU} \sim \frac{1}{\lambda},$$

$$\frac{1 - \exp(-(u+1)\hat{p}_{nU})}{\hat{p}_{nU}} \sim (1 - e^{-\frac{1}{\lambda}})\lambda \lg n;$$

so we have

$$\liminf_{n\to\infty} \frac{e_n}{\lg n} \geq \lambda\,(1-e^{-\frac{1}{\lambda}}).$$

Secondly, consider that this holds for any λ large enough, and that $\lambda(1-e^{-\frac{1}{\lambda}})$ comes arbitrarily close to 1 (for $\lambda\to\infty$). This means

$$\liminf_{n\to\infty} \frac{e_n}{\lg n} \geq 1.$$

One should guess that, after so merciless neglections, our estimate is far from being sharp. This is not the case.

To estimate e_n from above, we split the expression into three parts:

$$r = \sum_{k=0}^{u} \prod_{h=0}^{k-1} q_{nH} a_{nK} ,$$

$$s = \sum_{k=u+1}^{v} \prod_{h=0}^{k-1} q_{nH} a_{nK} ,$$

$$t = \prod_{h=0}^{v-1} q_{nH} \cdot \sum_{k=v+1}^{\infty} \prod_{h=v}^{k-1} q_{nH} a_{nK} ,$$

$$e_n = r + s + t.$$

The products within r and s and outside t are ≤ 1 (as products of probabilities).

In t, $\prod_{h=v}^{k-1} q_{nH}$ is $\leq q_{nV}^{k-v}$ (replacing all factors by the largest: the first one). Thus

$$e_n \leq \sum_{k=0}^{u} a_{nK} + \sum_{k=u+1}^{v} a_{nK} + \sum_{k=v+1}^{\infty} q_{nV}^{k-v} a_{nK}.$$

a_{nm} is increasing (since $a_{nm} = \frac{m}{n+1} p_{n+1,m} + 1$, and m and $p_{n+1,m}$ are increasing); so we replace the first two sums by their last term multiplied by the number of terms. a_{nm} is $\leq \frac{m}{n+1} + 1$ (since $p_{n+1,m} \leq 1$); so we replace a_{nK} in the third sum by $\frac{K}{n+1} + 1$. Thus

$$e_n \leq (u+1)a_{nU} + (v-u)a_{nV} + \sum_{k=v+1}^{\infty} q_{nV}^{k-v}\left(\frac{K}{n+1}+1\right).$$

The expression is not decreased if we replace a_{nm} by $\hat{a}_{nm} = \frac{\exp \frac{n}{m}}{\exp\frac{n}{m+1} - 1}$. Indeed, $a_{nm} = \frac{m}{n+1}(1+\frac{1}{m})^{n+1} p_{n+1,m+1}$; using the upper estimate $\hat{p}_{n+1,m+1} = \frac{\frac{n+1}{m+1}}{\exp\frac{n}{m+1} - 1}$ for $p_{n+1,m+1}$, and the fact that $1+\frac{1}{m} \leq e^{\frac{1}{m}}$, we get $\hat{a}_{nm}$.

Now we choose $u = \lceil \lg \frac{n+1}{\log n} \rceil = \lceil \lg(n+1) - \lg(\log n) \rceil$, $v = \lceil \lg(n+1) \rceil$.

Substituting u and v, we get for n growing:

$$a_{nU} \sim 1,$$

$$a_{nV} \sim \frac{e}{e-1}.$$

For the first two terms we have therefore bounds that are

$$\sim \lceil \lg(n+1) - \lg(\log n) \rceil + 1$$

$$+ \frac{e}{e-1}(\lceil \lg(n+1) \rceil - \lceil \lg(n+1) - \lg(\log n) \rceil)$$

$$\sim \lg n.$$

In the sum, $V \approx n+1$, so $q_{nV} \sim q_{n,n+1} = 1 - p_{n,n+1} \sim 1 - \frac{1}{e-1}$ $= \frac{e-2}{e-1} < \frac{7}{16}$. K is 2^k. So, for n large enough, the sum is

$$\leq \sum_{k=v+1}^{\infty} \left(\frac{7}{16}\right)^{k-v} \left(\frac{2^k}{n+1} + 1\right) = \sum_{k=1}^{\infty} \left(\frac{7}{16}\right)^{k} \left(\frac{2^{k+v}}{n+1} + 1\right).$$

2^v being $\leq n+1$, this sum is

$$\leq \sum_{k=1}^{\infty} \left(\left(\frac{7}{8}\right)^k + \left(\frac{7}{16}\right)^k\right) = \frac{70}{9}$$

which of course vanishes as compared with $\lg n$.

Taken together, this means

$$\limsup_{n\to\infty} \frac{e_n}{\lg n} \leq 1.$$

Combined with the previous lim inf relation, we obtain

$$\lim_{n\to\infty} \frac{e_n}{\lg n} = 1, \qquad \text{or } e_n \sim \lg n .$$

(For small $n \geq 2$, e_n is very well approximated by $\lg n$+const. or rather $\lg(n+1)$+const., where const. is very nearly 2; this approximation grows better as n increases, as long as computations go.)

With trial overhead, $w > 0$, we have $a_{nK} + w$ instead of a_{nK} in the formula, so we have to evaluate $\sum_{k=0}^{\infty} \prod_{h=0}^{k-1} q_{nH}$, the number of trials. For small n:

$$e_1^{(w)} = 1 + w,$$

$$e_2^{(w)} = 3.6888 + 2.6416w,$$

$$e_3^{(w)} = 3.8527 + 2.8822w,$$

$$e_4^{(w)} = 4.1609 + 3.1649w,$$

etc.

For large n, we can reiterate the previous deduction replacing a_{nK} by 1, and obtain the same result: $\lg n$. Whence

$$e_n^{(w)} \sim (1+w)\lg n.$$

So Ethernet has *not* the property of expected transmission start within a bounded amount of time for any number of claims. Of course, in any given system, the number of claims is bounded, in the worst case, by the number of stations; nowhere ever can a system *really* go out of bounds.

Impact of "truncation" is now obvious. Doubling limitation slightly decreases e_n if done above a certain limit, and heavily increases e_n below that limit; the limit depends on n. It can be computed from the above formulae; from Section 1 it is immediately clear that n is still on the wrong side whereas some value between n and $2n$ is already on the right side. In practice, we *must* have some limit (because the equipment, and consequently the number representation, is bounded), so any number safely greater than, say, twice the highest-conceivable number of concurring claims would do. (In existing systems it is usually chosen even higher.) Such a limitation improves upon all e_n with lower n; the improvement, however, is imperceptibly small.
- Retrial limitation bans "infinity" (by discarding stubborn cases as failures). Since probabilities can be computed, it seems reasonable to set it to a *very* low probability, and to indicate, not "retry later", but "call system designer".

ALTERNATIVES

So far we have only seen that Ethernet fails to be "best-conceivable" but there have been no arguments that it

ought to be really better, nor that anything better is really possible.

Incidentally we mention a method, sometimes advocated as "unfair but fast": every station has *its* slot to start in; collisions are excluded. Unfairness left apart, we have of course at least as many slots as stations, and if only a single station wants to transmit, we have to wait (provided it is any station, with equal probability) for a number of slots equal half the number of stations; while Ethernet would start in the very first slot in this case, and would expectedly start in lg of that number in the very worst of cases. With a heavy load, that "fast" system would become a private plaything of a born aristocracy, but even then it will perform presumably slower than Ethernet if not the King or the Dauphin personally is pulling the wire. - Rotating priority adds to complications and to fairness, not to speed.

Ethernet is not bad at all. For patriarchal applications like our LRI net it is better than needed. But there *are* cases where economy on slots can tangibly influence system efficiency. Imagine a 30 Mbit/sec cable of, say, a 1000 m length, involved in page transfer for memory management in behalf of a handful of small-and-fast processors, and handling, in addition, some buffered character peripherals. Slot time becomes comparable with transmission time (a page transfer takes but a few slots' time, a Chaosnet STS takes less than one). If some process control with very many very short records is involved, the situation is much more the same. Small differences in small-n handling would not be very perceivable in this case either. Yet mass collisions will arise, with a positive probability, although very rarely. (Very rarely since, in a system that functions at all, there are no more claims than successful transmissions, nor less of course; so we have one claim for each transmission interval, *if* claims need not be repeated.) In a jam, the same claims (except at most one of them) must be presented repeatedly. If this is accompanied by a significant increase of the average time consumed in servicing a single claim, this will have a cumulative effect, causing a slowly-melting freeze-in (while new claims may mushroom).

How to improve upon things?

Ethernet strategy has a double foible: its inconsiderate rush on the trough in case of trust, and its desperate yielding to force in case of frust. (Ethernet people insist that their strategy is "the most aggressive strategy" that can be adopted. They don't define what "aggressive" is to mean. If it means *this*, they are right.)

The problem of course is to make out the actual value of n, as reliably as possible. The strategy can be as good as our knowledge of n. No extreme accuracy is needed; to know n within a power of two is sufficient. We have shown that m can be practically anywhere in an interval $[x,2x]$ with n fixed; since the formula depends on $\frac{n}{m}$, the same is true for n with m fixed. One may wish for a tolerance not only in n but simultaneously in m - since choosing m a power of two has the advantage that random numbers come cheaper and faster. Our formulae show that

even this is compatible with d_{nm} well under 4. (The actual situation is much better. For $n \leq 16$, we find, by explicitly computing d_{nm}, that for each interval $2^{j-1} < n \leq 2^j$ there is a two's power value of m with $d_{nm} < 2.9$ in the worst case and still better on the average; m can be chosen 2^j. And for $n > 16$ all our formulae can be sharpened.) - Ethernet is wrong not in choosing m a two's power but in not choosing it in a better way.

FEEDBACK

To guess upon n, we have the state (and history) of the system. (As a matter of fact, Ethernet strategy itself is a - very ingenious - automatic feedback on this.)

We start with several simple qualitative remarks.

As we have an expectation for the transmission slot as a function of n and m, turning this round, we have an empirical estimate for n by knowing m and observing the slot. Each or all of the (observable) values of b and c, and so a, will do. The mathematical form (and even common sense) shows that they vary very much in parallel, so that a too early decision may be a warning even if successful, and a well-timed decision may be a reassurance even if a collision (a still commoner sense suggests the contrary).

After a successful transmission, there is no reason to think that all earlier claims have disappeared. Exactly one of them has. The simplest approach is to take *n* unchanged; or it can be diminished by one and increased by a quantity proportional to transmission *time*. (As futile as load *distribution* models seem to be, *overall* system load can be tolerably estimated.)

The strategy must be "stable" in the sense that newcomers can catch up, rules can be violated (with transient loss in efficiency).

Finally, we can't wire in a full handbook of probability theory. Some fairly simple automatism is needed to control the process.

Many such strategies are conceivable. We give a single specimen, a *very* primitive one; call it Othernet or Nethernet.

It comes in three versions.

V0

There is a current value of *m*, initially 1.
If after a transmission a pause of *m* slots elapsed, reset *m* to 1.
If an "event" - success or collision - occurred, call *h* the number of the slot it occurred in. (*h* is of course definite and detectable: if we are in the *m*-long aftermath of the previous event, slot count starts at detecting event termination; if it is later, take *h*=1.) Now if *h* is 1, double *m*, else halve *m*.

We need a "newcomer's rule" or "concord rule". Instead of sophisticated adaptivity, we suggest this.

Every transmission contains in its header the current value of *m* (or rather its exponent-of-two, so a few bits suffice). This is to be handled the same way as the "address", that is, it is heard by every receiver. Receiving stations give credit to it (whatever their previous belief).

Now the first ("creational") transmission is made with *m*=1. For newcomers, the "clean" rule would be: wait for the first foreign transmission (or as long as *m* is certain to be reset to 1). There is no need to do so. Anyone can butt in with any *m* he likes. If he gets through, *m* will be falsified. We'll see that in the worst case (going with *m*=1 when *m* should be high) we will need about Ethernet-time for the next action, and practically no loss for subsequent actions. So if intrusions are rare, they can be tolerated. Similarly, stations that have missed a foreign transmission (they *are* newcomers in our sense) may go on with any *m*, preferably their previous *m* (which will be seen the best value, and which is the simplest thing to do since it means not to do anything special).

This is not much more complicated than Ethernet. (We could even drop the resetting of *m*, which will be seen a loss, but a small loss, in efficiency.)

Intuitively, the gist of what happens can be grasped this way. Note that a_{nm}, the expected "event" slot, is in the "best-conceivable" case (take $m=1$ for $n=1$, $m=n+1$ for small $n\geq 2$, $m\sim n$ in the limit): $a_{11}=1$, $a_{23}=1.5556^{-}$, $a_{34}=1.5625$, $a_{45}=1.5664$, etc., going to $\frac{e}{e-1}=1.5820^{-}$. So the strategy could be very roughly characterized by saying that if the experienced event slot is smaller than what was to be expected in case of the best choice of m, we double m, if it is greater than that, we halve m.

V1

The same as V0 except: do not halve m to 1.

Intuitively, this aims at avoiding "hopeless trials" during contests.

V2

The same as V0 except that the doubling|halving rule is: double m for $h<2$, halve m for $h>2$, leave it unchanged for $h=2$.

Intuitively, this increases the number of slots and decreases the number of trials (to be used in case of higher values of the trial overhead w).

We estimate the performance.

We temporarily forget of the rule resetting m to 1 after a pause.

m is always a power of two, we denote it 2^j.

For V0, the probability of doubling (i.e., of $h=1$) is $1-(1-2^{-j})^n$ - use the arguments given while establishing a_{nm} -; the probability of halving is $(1-2^{-j})^n$. For V1, the same holds true except in case $j=1$ the probability of halving is 0; instead, j remains unchanged with probability 2^{-n}. For V2, doubling has again probability $1-(1-2^{-j})^n$, no change has $(1-2^{-j})^n-(1-2^{-j-1})^n$, and halving has $(1-2^{-j-1})^n$.

Departing from any initial j, or any initial distribution of j, we can find, step by step, the subsequent distributions of j.

Similarly, we can compute the expected outcome of a contest.

We write the transmission-slot formula in this shape:

$$\sum_{k=1}^{\infty} \sum_{j=0}^{\infty} r_{nkj} a_{n2^j} .$$

The formation rule is the following.

r_{n1j} are the "density" for the initial distribution. (They do not really depend on n.)

We denote by t_{njh} the probability of the transition $j\to h$, and by s_{njh} the product $t_{njh} q_{n2^j}$. (They depend of course on whether V0, V1 or V2 is used, and are 0 for $|j-h|>1$.)

Now

$$r_{n,k+1,j} = s_{n,j-1,j} r_{n,k,j-1}$$
$$+ s_{njj} r_{nkj}$$
$$+ s_{n,j+1,j} r_{n,k,j+1}.$$

(To see this, note that the inner sum is the contribution of the k^{th} trial. - Note, too, that the value for a mixture of distributions is the mixture of values, and that the order of summation can be changed by absolute convergence.)

The trial-number formula is obtained by leaving out the a_{n2j}.

For the "most general" case, replace the n by n_k.

We are going to examine the behaviour in the long run.

First we suppose that n remains unchanged (or but negligibly varying), not only during a contest but for a whole period of system activity.

There is no a priori guaranty that the distribution of j converge anywhere. Consider, however, that the process is an (infinite-state) Markov chain wherein (a) every state is reachable from every state in a finite number of transitions with positive probability, (b) there is no "leakage towards infinity", that is, for every $\varepsilon>0$ a j_ε can be given such that the sum of the probabilities of states with $j \le j_\varepsilon$ remains $\ge 1-\varepsilon$ after a sufficient

number of transitions. It follows from this that the chain is ergodic, and there exists a stationary distribution to which all distributions converge. (In the present case, convergence is fairly fast, as can be estimated from the fact that, for a_j far from the "good range", the probability of changing in the "good direction" is very high. But we are not going to go into this here.)

We denote the "density function" of j for the stationary distribution by $s_n(j)$.

For VO, we have the obvious "boundary condition" $s_n(j)=0$ for $j<0$; $s_n(0)$ is not 0, we denote it temporarily by a. For $j \ge 0$, the following must hold by stationarity:

$$s_n(j)=(1-(1-2^{-j+1})^n)s_n(j-1)+(1-2^{-j-1})^n s_n(j+1).$$

(This states simply the fact that s_n must be unchanged by transitions.)

Rewriting this, we get the "recursion"

$$s_n(j+1) = \frac{2^{(j+1)n}s_n(j)-(2^{(j+1)n}-(2^{j+1}-4)^n)s_n(j-1)}{(2^{j+1}-1)^n} .$$

Any constant multiple of a solution of this is obviously a solution. Conversely, any two solutions satisfying the boundary condition are constant multiples of each other. (Let $s_n^{(0)}$ and $s_n^{(1)}$ be two solutions, both 0 for $j=-1$, $s_n^{(0)}=a_0$ and $s_n^{(1)}=a_1$ for $j=1$. Then $\frac{1}{a_0}s_n^{(0)}-\frac{1}{a_1}s_n^{(1)}$ is identically 0 by the recursion.) The requirement that s_n be a density, and therefore $\sum_{j=0}^{\infty} s_n(j)=1$, fixes that constant multiplier, and makes $s_n(j)$ uniquely defined. (We know, *and* will find, that it *exists*.)

The recursion can be shortened to a one-step recursion:

$$s_n(0) = a,$$

$$s_n(j+1) = \frac{2^{(j+1)n} - (2^{j+1}-2)^n}{(2^{j+1}-1)^n} s_n(j), \quad j \geq 0.$$

(Substitution shows that this defines the same function.)

Let us denote

$$u_n(j) = \frac{2^{jn} - (2^j-2)^n}{(2^j-1)^n},$$

$$t_n(j) = \prod_{h=1}^{j} u_n(h).$$

Convening, as usual, that an empty product means 1, $t_n(0)$ is 1.

Then we must obviously have

$$a = 1 / \sum_{k=0}^{\infty} t_n(k),$$

and get

$$s_n(j) = t_n(j) / \sum_{k=0}^{\infty} t_n(k), \quad j \geq 0.$$

For V1, we have the boundary condition $s_n(j)=0$ for $j<1$, $s_n(1)=a\neq 0$. By stationarity, we have for $j=1$

$$s_n(1) = (\tfrac{1}{2})^n s_n(1) + (\tfrac{3}{4})^n s_n(2),$$

and for $j>2$ the same relation as with V0.

Repeating the argument, we happen to get the same shortened recursion as with V0 except it starts one later. Writing

$$t_n(j) = \prod_{h=2}^{j} u_n(h) = \prod_{h=1}^{j-1} u_n(h+1)$$

(with the same u_n as previously), we get

$$s_n(j) = t_n(j) / \sum_{k=1}^{\infty} t_n(k), \quad j \geq 1.$$

For V2, the boundary condition is the same as for V1. The stationarity condition yields now

$$s_n(j) = (1-(1-2^{-j+1})^n) s_n(j-1)$$
$$+ ((1-2^{-j})^n - (1-2^{-j-1})^n) s_n(j)$$
$$+ (1-2^j)^n s_n(j+1)$$

for $j \geq 1$.

This gives a two-step recursion, for which we again have a one-step recursion:

$$s_n(1) = a,$$

$$s_n(j+1) = \frac{2^{jn} - (2^j-1)^n}{(2^j-1)^n} s_n(j), \quad j \geq 1.$$

With

$$u_n(j) = \frac{2^{jn} - (2^j - 1)^n}{(2^j - 1)^n},$$

$$t_n(j) = \prod_{h=1}^{j-1} u_n(h)$$

we get

$$s_n(j) = t_n(j) / \sum_{k=1}^{\infty} t_n(j), \qquad j \geq 1.$$

Now we can compute, for small n, the expected transmission slot in case of stationary distribution. Denote them by $f_n^{(w)}$, $g_n^{(w)}$, $h_n^{(w)}$ for V0, V1, V2, resp.

n	$f_n^{(w)}$	$g_n^{(w)}$	$h_n^{(w)}$
1	2.0000+1.0000w	2.2654+1.0000w	2.5000+1.0000w
2	3.2190+1.5639w	3.1274+1.4099w	3.4466+1.3000w
3	3.2939+1.5183w	3.2601+1.4623w	3.6291+1.3041w
4	3.4047+1.5297w	3.3915+1.5095w	3.7627+1.3159w

etc.

We examine the behaviour for n large.

V0:

Denote by I_a the interval $[\lg n-a,\ \lg n+a]$, and by S_a the sum of all terms of the transmission-slot formula for the stationary distribution wherein some $m=2^j$ with $j \notin I_a$ occurs. We find that for any $\varepsilon>0$ an a can be chosen in such a way that $S_a<\varepsilon$ - namely, the $s_n(j)$ are so small for $j \notin I_a$ that everything multiplied by them becomes small enough, even added together. (This requires a trivial but tedious estimation of the terms, which we do not reproduce here. We just mention that for a seven decimal digits' precision we can confine us to the interval $[\lg n -5,\ \lg n+8]$.)
The same holds for the trial-number formula.

The fact of matter is that the midpoint of I_a depends on n but its length does not.

First we compute the density function. We start (say) at the left endpoint of I_a, $\lg n-a$, and take the least integer in the range, call it b. $0 \leq \lg n-a-b<1$. We assign an arbitrary value (say 1) to r_0. Then we compute

$$r_k \leftarrow \frac{2^{(b+k)n} - (2^{b+k}-2)^n}{(2^{b+k}-1)^n} r_{k-1} = u_n(b+k) r_{k-1}$$

for $k=1,2,\ldots,[2n]$; and

$$\bar{r} \leftarrow \sum_{k=0}^{[2n]} r_k .$$

(In practical computation, the interval is not symmetric around $\lg n$; the left wing is shorter than the right wing.)

Finally, we divide all r_k by $\bar{r}$, and take the results as approximations to $s_n(b)$, $s_n(b+1)$, etc.

This works for all n. For $n\to\infty$, we have

$$u_n(j)=(1+\frac{1}{2^j-1})^n-(1-\frac{1}{2^j-1})^n$$

$$\sim \exp(2^{-j}n)-\exp(-2^{-j}n)$$

$$= 2\sinh(2^{-j}n)$$

uniformly for all $j \in I_a$.

Writing $c=\lg n-b$ $(a-1<c\leq a)$:

$$u_n(b+k) \sim 2\sinh(2^{-c+k}).$$

So we can, for large n, approximate $s_n(j)$ by using these values in the computation of the r_k.

Similarly, for $j \in I_a$ and large n, we can use the asymptotic expressions for q_{nm} and a_{nm}, thus we have

$$q_{n2^{b+k}} \sim 1-\frac{2^{c-k}}{\exp(2^{c-k})-1} \text{ and } a_{n2^{b+k}} \sim \frac{\exp(2^{c-k})}{\exp(2^{c-k})-1};$$

in the transition probabilities, $(1-2^{-b-k})^n \sim \exp(2^{c-k})$.

All this is already independent of n except for the small variations of c. ($\lg n-c$ is an *integer*, so the fractional part of c depends on that of $\lg n$. - Only

the fractional part causes a problem: using $c+1$ instead of c is as if we had chosen $a+1$ instead of a, so this approximates the same values, with maybe greater precision.). These variations of c cause the remaining difficulty, which I cannot fully overcome.

The guess of course is that the limit does not depend on c. For $n\leq 25$ (until computations with the non-limit formulae have been made) $f_n^{(w)}$ shows no signs of oscillations. (It would be altogether sufficient to show that both f and $f_n^{(w)}$ are non-decreasing as n increases, which is intuitively rather plausible. It would be enough to show that oscillations fade away.) Moreover, computations of the limit formula have been made for different values of c between $a-1$ and a, partly equidistantly, partly at "Monte Carlo" points, giving all the same result, well within normal arithmetic error bounds (using an arithmetic providing a seven decimal digits' precision in the four species, the error is in the sixth digit, the number of elementary steps being several tenthousands). So all computations agree upon

$$\lim_{n\to\infty} f_n^{(w)} \approx 3.8631+1.5759w.$$

We show at least that the result (if not independent of c) can only boundedly differ from the above value for any c, and $f_n^{(w)}$ is bounded. To show this, it is sufficient to show that all involved quantities depend continuously on c in $a-1\leq c\leq a$. Of $q_{n2^{b+k}}$, $a_{n2^{b+k}}$ and the transition probabilities we know this, as well as of their asymptotic expressions (since they are continuous in m). We know nothing of $s_n(j)$ as it is not even defined yet for non-integer j.

We extend the definition of $s_n(j)$ to all real $j \geq 0$.

u_n is defined for all positive real values of the argument. Let us denote

$$\hat{t}_n(j,x) = 2^{xn} \prod_{h=1}^{j} u_n(h+x), \quad 0 \leq x \leq 1.$$

We have

$$\hat{t}_n(j,0) = t_n(j), \ \hat{t}_n(j,1) = t_n(j+1).$$

We now *define* to be

$$t_n(j) = \hat{t}_n([j], j-[j]) \text{ for all real } j \geq 0.$$

This implies the corresponding definition for $s_n(j)$.

$t_n(j)$ and $s_n(j)$ are continuous. For integer j, they have their original values. Moreover, the relation

$$\frac{s_n(j+1)}{s_n(j)} = \frac{t_n(j+1)}{t_n(j)} = u_n(j+1)$$

remains valid for non-integer j.

Now compute the r_n, starting (instead from b) from any value b^* near a. This amounts to approximate $s_n(j-(b-b^*))$. The approximation is obviously continuous in b^*, thus in c; s_n itself is continuous by the above. For $j \in I_a$ and $n \to \infty$, the approximations converge uniformly, so there exists a limiting function, itself continuous,

call it $s(j)$. In addition, for $j \in I_a$ and n large enough, $|s_n(j)-s(j)|$ is less than (say) 2ε.

V1:

We could recapitulate, with minor differences, the above argument. We replace this by the remark that V1 behaves for $n \to \infty$ *evidently* the same way as V0.

V2:

The argument runs analogously, the only not-absolutely-obvious difference being

$$\frac{s_n(j)}{s_n(j-1)} = \frac{t_n(j)}{t_n(j-1)} = u_n(j-1) = (1-2^{-j+1})^{-n}-1$$

$$\sim \exp(2^{-j+1}n)$$

$$= \exp(2^{-c+k+1})-1.$$

The "agreed upon" value in the limit is

$$\lim_{n\to\infty} h_n^{(w)} \approx 4.3000+1.3546w.$$

(The numerical behaviour is the same as with V0. We forge an additional intuitive argument in favour of the independence of $c-[c]$: We have had altogether four different expressions, all displaying numerical insensitivity to c. Don't tell it is chance.)

In the extension of $s_n(j)$ to non-integer j we use here

$$\hat{t}_n(j,x) = (2^{x(n-1)+1}-1) \prod_{h=1}^{j-1} u_n(h+x).$$

Now suppose n undergoes violent changes. (Violent changes are only possible upwards; decrease is always one by one.)

The worst case is if we are at m=1 when the burst of n comes. We denote the "worst-burst" expectations by $\hat{f}_n^{(w)}$, $\hat{g}_n^{(w)}$, $\hat{h}_n^{(w)}$.

For small "bursts":

n	$\hat{f}_n^{(w)}$	$\hat{g}_n^{(w)}$	$\hat{h}_n^{(w)}$
1	1.0000+1.0000w	1.0000+1.0000w	1.0000+1.0000w
2	3.7079+2.8823w	3.5558+2.7302w	3.5600+2.7163w
3	3.8464+3.0902w	3.7534+2.9973w	3.7452+2.9630w
4	4.1495+3.3767w	4.0944+3.3216w	4.0595+3.2569w

etc.

For $n\to\infty$, all three are easily shown $\sim(1+w)\lg n$, the same as Ethernet. (Hint: For steps with $m=2^j$ where $j\leq \lg n - \lg(\log n)$, doubling probabilities go to doubling certainties.)

If we are at m=2, values for n=1 increase by 0.5, those for $n\geq 2$ decrease by $1+w$; etc. Generally, if we are at m, are faced with n, and $n-m$ is large, the expected time will be near $(1+w)(\lg n - \lg m)$. (Hint: Follow the previous hint.)

So we may be better off than Ethernet even in case of a burst, since we may happen to be at a better value of m than 1. But this is not the point. The essential difference can be assessed by considering transmissions *following* a burst. While Ethernet keeps starting at m=1,

the doubling/halving strategies climb to the "good range" and remain around. (So they have their "lg n" overhead only once, whereas Ethernet, even if no additional claims arise, continues with $\lg(n-1)$, $\lg(n-2)$, ..., which adds up to $\sim n\lg n$.) There is no "iceberg".

It is clear, not only that the size of bursts is limited by the number of stations, but that n^*-bursts - that is, those with $n_{k+1}-n_k=n^*$ - cannot occur more frequently on the average than once per n^* transmissions because, as servicing is one by one, the number of stations would be exceeded otherwise. Note that the argument depended on the existence of a bound on the number of concurring claims but not on its size. It follows that long contest times are rare enough to provide a bounded average for *any* distribution of the n; this average is always (much) less than $M+1+w$ where M is the upper limit of the expectation in the stationary or in the decreasing-by-one case of V0, V1 or V2. (Hint: Estimate the contribution of bursts to the average by inspecting "small bursts" directly, and considering for large bursts the fact that $\frac{\lg n_{k+1}-\lg n_k}{n_{k+1}-n_k} \leq \frac{\lg(n_{k+1}-n_k)}{n_{k+1}-n_k} = \frac{\lg n^*}{n^*}$ is $\leq\frac{1}{2}$ and is decreasing when n^* increases.)

(If we *had* a theory on the distribution, e.g., that new claims appearing per time interval are binomial or Poisson, we could go farther and give sharper estimates.)

Now we resume the "reset m to 1" rule.

Only the state after a transmission with n=1 is affected. (If n has been ≥ 2, at least $n-1$ claims present themselves in the "m-long aftermath", and there will be no "pause".)

So resetting depends on whether the first claim after a single-claim servicing appears after a pause or earlier. As we refrain from making assumptions about the distribution of claims, we cannot state anything precise. All we know is that the resetting rule makes the case $n=1$ somewhat better (as resetting *can* occur) without perceivably deteriorating cases $n\geq 2$ (as in case of a pause claims could not have been cumulated, so the next competitors will be only claims raised within a single slot's time, the number of which being with a very low probability higher than 1, *whatever* the distribution of claims is). Nevertheless, the case $n=1$ remains worse than with Ethernet. It could be argued that frequent $n=1$ states are likely to occur in small-load periods, and we might accept things as they are. Or else the following two rules could be added:

A. Don't double m if it has been 1 and transfer succeeded.

B. Halve m if it has been 2 and no collision occurred.

Rule A makes a sequence with $n=1$ go ultimately to $m=1$ with probability 1; only the case $n=1$ is affected. Rule B makes this process fast; cases $n\geq 2$ are also slightly affected (especially, if $n=2$, $m=2$ is followed by $n=1$, this is an improvement; if it is followed by $n\geq 2$, this is a degradation). - Even so, the case $n=1$ is worse than with Ethernet. (It *tends* to $m=1$ but, e.g., after an $n=2$, $m=2$, the probability of getting $m=1$ for the next transmission is only $\frac{1}{2}$, and certainty of $m=1$ starts only for the transmission thereafter; Ethernet has *always* $m=1$.) Yet this might already pass as "comparable".

Summing up: Comparison of the V strategies with Ethernet is clear. Comparing them with each other, we see that the additional rules make the case $n=1$ fairly the same for all of them, and, for all n, V1 is generally preferable to V0 (but more complicated), V1 is preferable to V2 if $w<2$, V2 to V1 if $w>2.9$ in the stationary case, and V2 is generally better than V1 for bursts. If *one* of them has to be chosen, choose V2. - Fortunately, we need no risky assumptions on the distribution of claims to compare the methods.

"Truncation" can be handled by the same techniques (the discussion becomes conceptually simpler still, as we have here a finite-state Markov chain). As to its effects, precisely the same can be said as in the case of Ethernet.

Note that some kind of "priority" can be introduced by deliberately biasing the lot-slots; not of course without detriment to overall performance. Nevertheless, rare urgent transmissions ("emergency messages") may safely be started by forcing (or only using) $m=1$.

- * -

Such methods are unable to *reach* the "best-conceivable" performance. First, all our estimates have a *variance*, so false guesses are a mathematical necessity. Secondly, we use the *prevailing* state in handling the *next* action, so state changes keep confuting premises.

All this suggests the following approach.

FEEDFORWARD

If we sacrifice the first slot, during which the claimants put a *mark* of their wish on the line wherefrom their *number* can be estimated, then we can approximate the "best-conceivable" values with the wasted slot added. If this can be arranged in a way that either the marking does not take a full slot or a claimant that happened to be alone is immediately started (that is, the "mark" is at the same time the beginning of the message), then we are still better off.

Recall that to know n within a power of two is sufficient.

The author is unfortunately not a hardware man, so his/her ideas are prone to be amateurish. Two of them are brought forward, not in earnest, merely to indicate ways of thinking.

(a) Put some voltage on the line. The questions are of course: what differences are possible, and how long it takes to measure.

(b) Have a "minimal distance" between stations, enough to transmit one bit. (This restricts the applicability to fast lines, but only those are of interest here.) At transmission termination, each claimant appends a bit to the tail. These bits can be counted everywhere as they gather. Coming from two directions, at most half of them is lost by overlapping arrival. (The station just terminating is in the worst position, especially if located in the middle of the line.) - On very fast lines, stations can perhaps be interspaced in a way that no overlaps arise.

APPENDIX: NUMBERS

The following quantities are tabulated:

- The values $d^{(w)}_{n,n+1}$. (With w=0, the tabulated values are best-possible for $n \geq 2$; for n=1, $d^{(w)}_{11}=1+w$ is better.)

- The values $d^{(w)}_{n,2n+2}$.

- The values $d^{(w)}_{nh}$ with h the two's power greater than $\frac{2}{3}(n+1)$ and not greater than $\frac{4}{3}(n+1)$. This sequence has no limit as $n \to \infty$; writing the values as $a_n + b_n w$, both a_n and b_n fluctuate, but we have liminf $a_n \geq 2.7182$, limsup $a_n \leq 2.9878$, liminf $b_n \geq 1.4893$, limsup $b_n \leq 2.3212$. (For some n, an unambiguously better two's-power m exists; e.g., $d^{(w)}_{11}$ is $1+w$, $d^{(w)}_{22,32}$ is $2.8088+1.4417w$.)

- The Ethernet expectations: $e^{(w)}_n$.

- The "worst burst" cases, that is, the expectations when starting with m=1, for the strategies V0, V1, V2: $\hat{f}^{(w)}_n$, $\hat{g}^{(w)}_n$, $\hat{h}^{(w)}_n$.

- The "stationary" cases, that is, the expectations with unvaried n, for V0, V1, V2: $f^{(w)}_n$, $g^{(w)}_n$, $h^{(w)}_n$.

Tabulations go from n=1 to n=25; the last row shows the behaviour as $n \to \infty$. Values are rounded to four decimal places.

n	$d^{(w)}_{n,n+1}$	$d^{(w)}_{n,2n+2}$	$d^{(w)}_{nh}$	$e^{(w)}_{n}$
1	1.5000+1.0000w	2.5000+1.0000w	1.5000+1.0000w	1.0000+1.0000w
2	2.3333+1.5000w	3.0333+1.2000w	2.5000+1.3333w	3.6888+2.6416w
3	2.3810+1.5238w	3.0857+1.2190w	2.3810+1.5238w	3.8527+2.8822w
4	2.4475+1.5625w	3.1275+1.2346w	2.4583+1.7778w	4.1609+3.1649w
5	2.4925+1.5886w	3.1556+1.2450w	2.6423+1.4015w	4.4622+3.4302w
6	2.5247+1.6071w	3.1757+1.2524w	2.5681+1.5062w	4.7242+3.6629w
7	2.5488+1.6210w	3.1908+1.2580w	2.5488+1.6210w	4.9467+3.8638w
8	2.5676+1.6318w	3.2026+1.2624w	2.5707+1.7472w	5.1370+4.0385w
9	2.5826+1.6405w	3.2120+1.2659w	2.6260+1.8860w	5.3030+4.1927w
10	2.5949+1.6476w	3.2198+1.2687w	2.7106+2.0390w	5.4506+4.3311w
11	2.6052+1.6535w	3.2262+1.2711w	2.7321+1.4453w	5.5843+4.4570w
12	2.6139+1.6584w	3.2317+1.2731w	2.6850+1.4976w	5.7070+4.5730w
13	2.6213+1.6627w	3.2363+1.2749w	2.6546+1.5524w	5.8204+4.6806w
14	2.6278+1.6664w	3.2404+1.2764w	2.6380+1.6097w	5.9259+4.7810w
15	2.6334+1.6697w	3.2440+1.2777w	2.6334+1.6697w	6.0245+4.8750w
16	2.6384+1.6725w	3.2471+1.2788w	2.6392+1.7325w	6.1170+4.9634w
17	2.6428+1.6751w	3.2499+1.2799w	2.6543+1.7982w	6.2040+5.0468w
18	2.6468+1.6774w	3.2524+1.2808w	2.6779+1.8671w	6.2859+5.1256w
19	2.6504+1.6794w	3.2546+1.2816w	2.7093+1.9393w	6.3635+5.2002w
20	2.6536+1.6813w	3.2567+1.2824w	2.7480+2.0150w	6.4369+5.2711w
21	2.6565+1.6829w	3.2585+1.2831w	2.7938+2.0942w	6.5068+5.3387w
22	2.6592+1.6845w	3.2602+1.2837w	2.8462+2.1774w	6.5733+5.4031w
23	2.6617+1.6859w	3.2617+1.2843w	2.7773+1.4673w	6.6369+5.4648w
24	2.6640+1.6872w	3.2632+1.2848w	2.7509+1.4935w	6.6977+5.5239w
25	2.6660+1.6884w	3.2645+1.2853w	2.7292+1.5202w	6.7562+5.5807w
lim	2.7183+1.7183w	3.2974+1.2974w		lg n(1+w)

n	$\hat{f}^{(w)}_{n}$	$\hat{g}^{(w)}_{n}$	$\hat{h}^{(w)}_{n}$	$f^{(w)}_{n}$	$g^{(w)}_{n}$	$h^{(w)}_{n}$
1	1.0000+1.0000w	1.0000+1.0000w	1.0000+1.0000w	2.0000+1.0000w	2.2654+1.0000w	2.5000+1.0000w
2	3.7079+2.8823w	3.5558+2.7302w	3.5600+2.7163w	3.2190+1.5639w	3.1274+1.4099w	3.4466+1.3000w
3	3.8464+3.0902w	3.7534+2.9973w	3.7452+2.9630w	3.2939+1.5183w	3.2601+1.4623w	3.6291+1.3041w
4	4.1495+3.3767w	4.0944+3.3216w	4.0595+3.2569w	3.4047+1.5297w	3.3915+1.5095w	3.7627+1.3159w
5	4.4537+3.6526w	4.4233+3.6222w	4.3609+3.5295w	3.4845+1.5410w	3.4797+1.5340w	3.8524+1.3238w
6	4.7195+3.8949w	4.7035+3.8788w	4.6222+3.7673w	3.5399+1.5475w	3.5382+1.5452w	3.9166+1.3291w
7	4.9443+4.1026w	4.9361+4.0944w	4.8449+3.9725w	3.5804+1.5515w	3.5798+1.5508w	3.9647+1.3328w
8	5.1357+4.2817w	5.1316+4.2776w	5.0360+4.1508w	3.6117+1.5544w	3.6116+1.5541w	4.0020+1.3355w
9	5.3022+4.4390w	5.3002+4.4369w	5.2027+4.3079w	3.6370+1.5567w	3.6369+1.5567w	4.0317+1.3376w
10	5.4504+4.5798w	5.4494+4.5787w	5.3510+4.4486w	3.6577+1.5588w	3.6577+1.5588w	4.0561+1.3393w
11	5.5846+4.7078w	5.5841+4.7073w	5.4850+4.5763w	3.6751+1.5605w	3.6751+1.5605w	4.0765+1.3407w
12	5.7077+4.8256w	5.7075+4.8254w	5.6077+4.6936w	3.6897+1.5618w	3.6897+1.5618w	4.0937+1.3419w
13	5.8216+4.9349w	5.8215+4.9347w	5.7212+4.8023w	3.7021+1.5629w	3.7021+1.5629w	4.1085+1.3429w
14	5.9276+5.0367w	5.9275+5.0366w	5.8267+4.9037w	3.7129+1.5639w	3.7129+1.5639w	4.1213+1.3437w
15	6.0266+5.1320w	6.0265+5.1320w	5.9254+4.9986w	3.7222+1.5646w	3.7222+1.5646w	4.1325+1.3445w
16	6.1192+5.2215w	6.1192+5.2215w	6.0179+5.0879w	3.7305+1.5653w	3.7305+1.5653w	4.1423+1.3451w
17	6.2064+5.3057w	6.2064+5.3057w	6.1050+5.1720w	3.7379+1.5659w	3.7379+1.5659w	4.1511+1.3456w
18	6.2885+5.3853w	6.2885+5.3853w	6.1871+5.2514w	3.7445+1.5664w	3.7445+1.5664w	4.1589+1.3461w
19	6.3661+5.4606w	6.3661+5.4606w	6.2648+5.3267w	3.7504+1.5669w	3.7504+1.5669w	4.1660+1.3466w
20	6.4397+5.5322w	6.4397+5.5322w	6.3384+5.3982w	3.7558+1.5674w	3.7558+1.5674w	4.1723+1.3470w
21	6.5096+5.6003w	6.5093+5.6003w	6.4084+5.4663w	3.7607+1.5678w	3.7607+1.5678w	4.1781+1.3473w
22	6.5763+5.6653w	6.5763+5.6653w	6.4750+5.5312w	3.7652+1.5682w	3.7652+1.5682w	4.1834+1.3477w
23	6.6400+5.7274w	6.6400+5.7274w	6.5386+5.5933w	3.7693+1.5686w	3.7693+1.5686w	4.1883+1.3480w
24	6.7008+5.7869w	6.7008+5.7869w	6.5994+5.6527w	3.7731+1.5689w	3.7731+1.5689w	4.1928+1.3483w
25	6.7594+5.8442w	6.7594+5.8442w	6.6579+5.7098w	3.7766+1.5692w	3.7766+1.5692w	4.1969+1.3485w
lim	lg n(1+w)	lg n(1+w)	lg n(1+w)	3.8631+1.5759w	3.8631+1.5759w	4.3000+1.3546w

OTHER TITLES IN THE SERIES:

KALAP, a Batch Monitor System, 1978.

MOTOR, a Mini-Micro Blow-Up, 1979.

KIBE, Character Handling, 1980.

Floating-Point Algorithms, as used at the Dept. for Automation in Engineering, 1982. (First printing: 1979.)

IN PREPARATION:

TALMUD, an Interpretive Technique, and KETHAB, a Way of writing Code for It. (A streamlining of the 1973 version.)

LÓHALÁL, a Chaotic Quicknet.

LASSO, or Quick Khachiyan.

Hash-Cash.

OTHER TITLES IN THE SERIES:

KALAP, a Batch Monitor System, 1978.

KIBE, Character Handling, 1980.

MOTOR, a Mini-Micro Blow-Up, (second printing) 1983.

LÓHALÁL, a Chaotic Quicknet, (second printing) 1986.

Floating-Point Algorithms, as used at the Dept. for Automation in Engineering, (third printing) 1985.

IN PREPARATION:

KOTOR, a File-Server System.

TALMUD, an Interpretive Technique, and KETHAB, a Way of writing Code for It. (A streamlining of the 1973 version.)

LASSO, or Qick Khachiyan.

Hash-Cash.