Probabilistic Methods for Algorithms and Stochastic Networks

Lecture 1

Philippe Robert

INRIA Paris-Rocquencourt

General Plan of Talks

Algorithms

— Access Protocols

— Congestion Control

— Divide and Conquer Algorithms

— Polling Systems

General Plan of Talks

Algorithms

— Access Protocols

— Congestion Control

— Divide and Conquer Algorithms

— Polling Systems

Generic Algorithms

General Plan of Talks

Mathematical tools

— Functional Transforms

— Random Walks

— Scalings of Stochastic Processes

General Plan of Talks

Mathematical tools

— Functional Transforms

— Random Walks

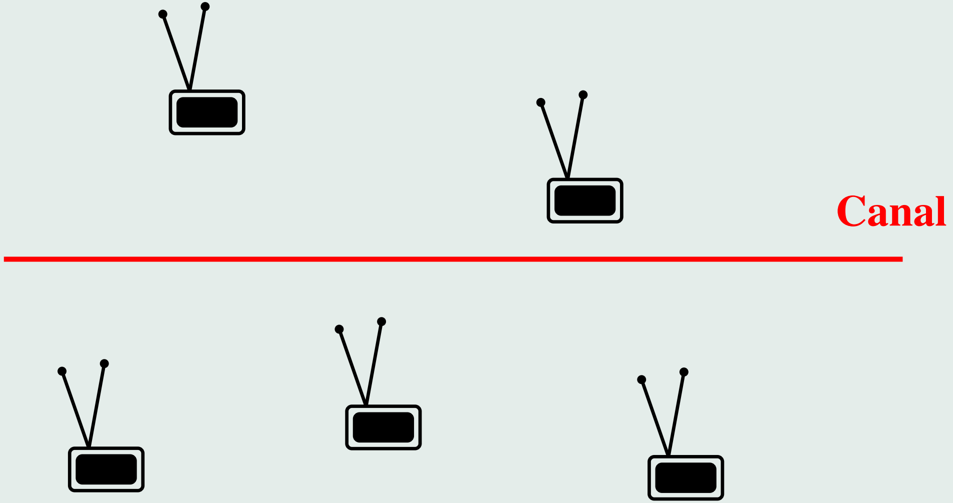
— Scalings of Stochastic Processes

Non-standard phenomena

— Non-Convergence in distribution.

— Oscillations

Problem of Access to a shared resource


Canal

Framework: A Distributed System

— N stations (transmitters) scattered.
No Central Control

— One communication channel.
Shared Resource

— ≥ 2 simultaneous attempts to transmit $\Rightarrow$ failure. Competition for Resource.

Framework: A Distributed System

— N stations (transmitters) scattered.
No Central Control

— One communication channel.
Shared Resource

— ≥ 2 simultaneous attempts to transmit $\Rightarrow$ failure. Competition for Resource.

Problem :Find a Decentralized Policy.

A Long Story

— Aloha (1968)
Abramson (Hawai)

— Ethernet (1973)
Metcalf and Boggs (Harvard)

— Cambridge Ring (1974)
Cambridge University

— Tree Algorithms (1979)
Capetanakis (MIT)
Tsybakov and Mikhailov (Acad. Sc.Moscow).

Information Available at a Station

Every time unit
A station can listen the channel and detect:

0 —Silence
no transmission attempt on the channel.

1 —Success
only one attempt.

2 —Collision
at least two attempts.

Information Available at a Station

Every time unit
A station can listen the channel and detect:

0 —Silence
no transmission attempt on the channel.

1 —Success
only one attempt.

2 —Collision
at least two attempts.

The channel conveys a ternary information.

Ethernet

Algorithm

Metcalf (Harvard) 1973

Each stations has a variable "counter" C.

— Arrival in the network: $C = 0$;

— After each failure to transmit $C \to C + 1$.

Algorithm

Station with counter equal to k:

Coin tossing

— Head (proba $1/2^k$): Transmission Attempt.

Algorithm

Station with counter equal to k:

Coin tossing

— **Head** (proba $1/2^k$): Transmission Attempt.

— **Tail** (proba $1 - 1/2^k$):
No transmission attempt.

Algorithm

k failures: transmission with proba $1/2^k$:

Algorithm

k failures: transmission with proba $1/2^k$:

— "new" messages are favored.

Algorithm

k failures: transmission with proba $1/2^k$:

— "new" messages are favored.

— Repeated collisions less likely.

A Markov chain on $\mathbb{N}$

$(b(n))$: State of the counter of a station at n

— $b(0) = 0$

— always experiences transmission failures.

A Markov chain on $\mathbb{N}$

$(b(n))$: State of the counter of a station at n

— $b(0) = 0$

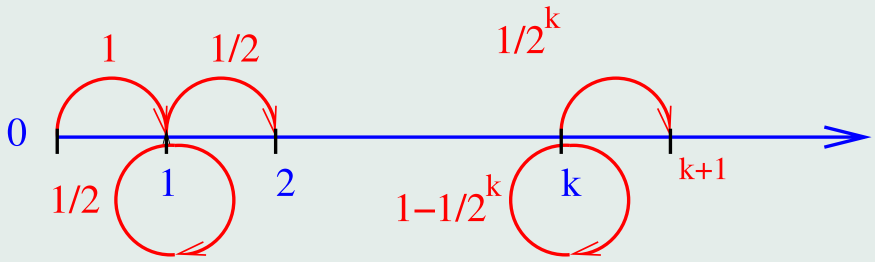
— always experiences transmission failures.

Transitions

— $\mathbb{P}(b(n) = k + 1 \mid b(n-1) = k) = \frac{1}{2^k}$

— $\mathbb{P}(b(n) = k \mid b(n-1) = k) = 1 - \frac{1}{2^k}$

A Markov chain on $\mathbb{N}$



A relation with counting algorithms

$Y_1, \dots, Y_m$ a sequence of elements of $\mathcal{S}$

Problem: what is the cardinality of

$$\{Y_1, \dots, Y_m\}?$$

A relation with counting algorithms

$Y_1,\ldots,Y_m$ a sequence of elements of $\mathcal{S}$

Problem: what is the cardinality of

$$\{Y_1,\ldots,Y_m\}?$$

Constraints

— One-line Algorithm

— Minimal memory and
number of operations.

Hash Function

$h : \mathcal{S} \longrightarrow [0,1]$;

Hypothesis

— $h(x)$ uniform on $[0,1]$.

— If $x \neq y$, $h(x)$ and $h(y)$ are independent.

In particular $h(x) \neq h(y)$.

Hash Function

If $\mathbf{card}\{Y_1, \ldots, Y_m\} = n = \mathbf{card}\{h(Y_1), \ldots, h(Y_m)\}$

It can be assumed that

— $Y_1, \ldots, Y_m$ are elements of $[0, 1]$

— $\{Y_1, \ldots, Y_m\} = \{U_1, \ldots, U_n\}$ where (U_i) are i.i.d. uniform on $[0, 1]$.

A Counter

— $C_0 = 0$.

— Step p:
If $Y_{p+1} < 1/2^{C_p}$ then $C_{p+1} = C_p + 1$.

C_p is a record value after inspecting p elements of the set $\{Y_1, \ldots, Y_p\}$

A Counter

The value of the counter C_m is the same as for n independent uniform random variables U_1,. . . , U_n on $[0, 1]$.

$C_m = D_n$ with

— $D_0 = 0$.

— Step p:
If $U_{p+1} < 1/2^{D_p}$ then $D_{p+1} = D_p + 1$.

(D_n) is our Markov chain $(b(n))$!

A Counting Algorithm

Since $D_n \sim \log_2 n$

$\Rightarrow$ Estimation of n with D_n.

A Counting Algorithm

Since $D_n \sim \log_2 n$
$\Rightarrow$ Estimation of n with D_n.

More details in Flajolet & Martin (1985), Flajolet (1985).

Growth of the Markov chain $(b(n))$

Proposition. For $n \geq 0$,

$$\frac{\log n}{\log 2} + \frac{\log \log 2}{\log 2} \leq \mathbb{E}(b(n)) \leq \frac{\log(n+1)}{\log 2}.$$

Growth of the Markov chain $(b(n))$

Proposition. For $n \geq 0$,

$$\frac{\log n}{\log 2} + \frac{\log \log 2}{\log 2} \leq \mathbb{E}(b(n)) \leq \frac{\log(n+1)}{\log 2}.$$

$$\lim_{n \to +\infty} \frac{\mathbb{E}(b(n))}{\log_2 n} = 1.$$

Growth of the Markov chain $(b(n))$

Proposition. For $n \geq 0$,

$$\frac{\log n}{\log 2} + \frac{\log \log 2}{\log 2} \leq \mathbb{E}(b(n)) \leq \frac{\log(n+1)}{\log 2}.$$

$$\lim_{n \to +\infty} \frac{\mathbb{E}(b(n))}{\log_2 n} = 1.$$

Convergence in distribution of $(b(n) - \lfloor \log_2 n \rfloor)$?

Geometric Random Variables

— G_k geom. r.v. with parameter $1 - 1/2^k$

$$\mathbb{P}(G_k \geq x) = (1 - 1/2^k)^{x-1}, \quad x \geq 1$$

Geometric Random Variables

— G_k geom. r.v. with parameter $1 - 1/2^k$

$$\mathbb{P}(G_k \geq x) = (1 - 1/2^k)^{x-1}, \qquad x \geq 1$$

$$G_k/2^k \sim E_k,$$

where E_k exp. r.v. with parameter 1.

Geometric Random Variables

$$\frac{1}{2^x}\sum_{k=0}^{x} G_k \sim H,$$

where

$$H = \sum_{k=0}^{+\infty} \frac{E_k}{2^k}$$

(E_k) i.i.d. exp. r.v. with rate 1.

Geometric Random Variables

$$\frac{1}{2^x}\sum_{k=0}^{x} G_k \sim H,$$

where

$$H = \sum_{k=0}^{+\infty} \frac{E_k}{2^k}$$

(E_k) i.i.d. exp. r.v. with rate 1.

H plays also an important role
for congestion control algorithm

Geometric Random Variables

Hitting time of n by $(b(n))$ has the same distribution as $G_0 + G_1 + G_2 + \cdots + G_{n-1}$

$$\mathbb{P}(b(n) \leq x) = \mathbb{P}\left(\sum_{k=0}^{x-1} G_k \geq n\right)$$

A convergence “result”

For $y \geq 0$

$$\mathbb{P}(b(n) - \lfloor \log_2 n \rfloor \leq y) \sim \mathbb{P}\left(H \geq 2^{-y} 2^{\log_2 n - \lfloor \log_2 n \rfloor}\right)$$

A convergence “result”

For $y \geq 0$

$$\mathbb{P}(b(n) - \lfloor \log_2 n \rfloor \leq y) \sim \mathbb{P}\left(H \geq 2^{-y} 2^{\log_2 n - \lfloor \log_2 n \rfloor}\right)$$

For n large

$$\mathbb{P}(b(n) - \lfloor \log_2 n \rfloor \leq y) \sim \mathbb{P}\left(H \geq 2^{-y} F(\log_2 n)\right)$$

where g is the periodic function

$$g(x) = 2^{x - \lfloor x \rfloor}.$$

A convergence “result”

For $y \geq 0$

$$\mathbb{P}(b(n) - \lfloor \log_2 n \rfloor \leq y) \sim \mathbb{P}\left(H \geq 2^{-y} 2^{\log_2 n - \lfloor \log_2 n \rfloor}\right)$$

For n large

$$\mathbb{P}(b(n) - \lfloor \log_2 n \rfloor \leq y) \sim \mathbb{P}\left(H \geq 2^{-y} F(\log_2 n)\right)$$

where g is the periodic function

$$g(x) = 2^{x - \lfloor x \rfloor}.$$

No CV in distribution. Flajolet (1985).

Back to Ethernet Algorithm

- Initial State: empty.
- Every time unit: Poisson arrivals (λ).
- Assumption: systematic failures
  External Source.

Back to Ethernet Algorithm

$\boldsymbol{X_i(n)}$: number of stations with counter $\boldsymbol{i}$ at time $\boldsymbol{n}$.

Proposition. $(X_i(n), i \geq 0)$ independent Poisson r.v. $X_i(n)$ has parameter

$$\lambda \sum_{k=0}^{n} \mathbb{P}(b(k) = i)$$

Proposition. $(X_i(n), i \geq 0$ independent Poisson r.v. $X_i(n)$ has parameter

$$\lambda \sum_{k=0}^{n} \mathbb{P}(b(k) = i)$$

Ingredient: If Y Poisson μ and (B_i) Bernoulli with parameter p then

$$\left(\sum_{i=1}^{Y} B_i, \sum_{i=1}^{Y} (1 - B_i) \right)$$

are independent Poisson r.v. with resp parameter $p\mu$, $(1 - p)\mu$.

Back to Ethernet Algorithm

Number of attempts A_n at time n is Poisson with parameter

$$\lambda \mathbb{E}(b(n+1)).$$

Back to Ethernet Algorithm

Number of attempts A_n at time n is Poisson with parameter

$$\lambda \mathbb{E}(b(n+1)).$$

$$P(A_n = 1) = \lambda \mathbb{E}(b(n+1)) e^{-\lambda \mathbb{E}(b(n+1))} \sim \lambda \frac{\log_2(n)}{e^{\lambda \log_2(n)}} = \lambda \frac{\log_2(n)}{n^{\lambda / \log(2)}}$$

Back to Ethernet Algorithm

Consequence **If** $\lambda > \log 2$

$$\sum_{n=1}^{+\infty} P\left(A_n = 1\right) < +\infty.$$

Back to Ethernet Algorithm

Consequence If $\lambda > \log 2$

$$\sum_{n=1}^{+\infty} P(A_n = 1) < +\infty.$$

Almost surely no transmission success after some finite time even without external source.

Ehernet unstable when $\lambda > \log 2$. Kelly (1985).

Back to Ethernet Algorithm

Aldous (1987) Ethernet unstable when $\lambda > 0$.

Back to Ethernet Algorithm

Aldous (1987) Ethernet unstable when $\lambda > 0$.

No simple proof yet !

Back to Ethernet Algorithm

Aldous (1987) Ethernet unstable when $\lambda > 0$.

No simple proof yet !

For $\lambda < \log 2$ an infinite number of successful transmissions.

The End for Today

Tomorrow: **Trees and Recursive Equations.**