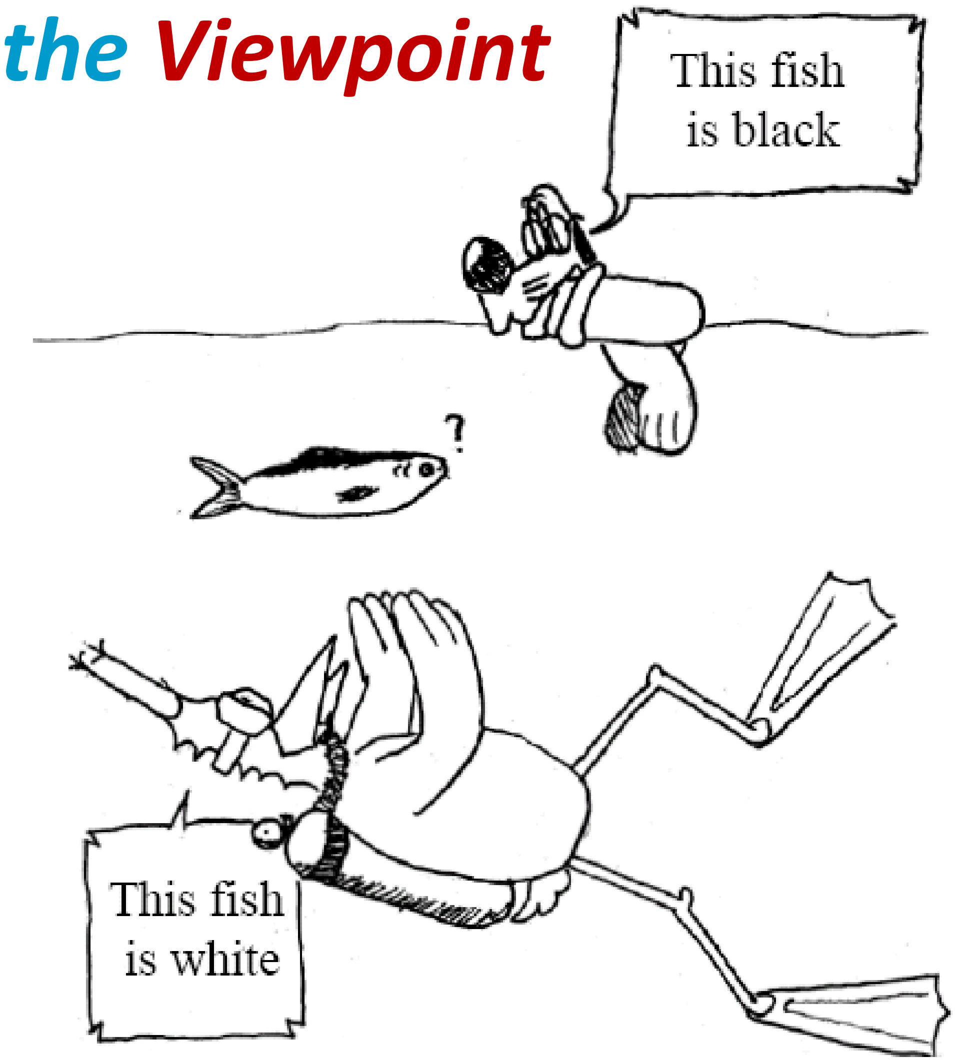
Palm Calculus
Made Easy
The Importance of the Viewpoint



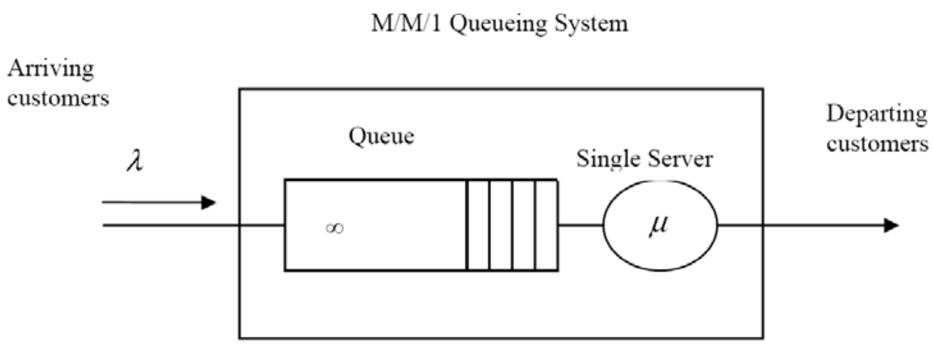
Contents

This is a branch of probability that is not well known (why?), though it is quite important for any measurement study.

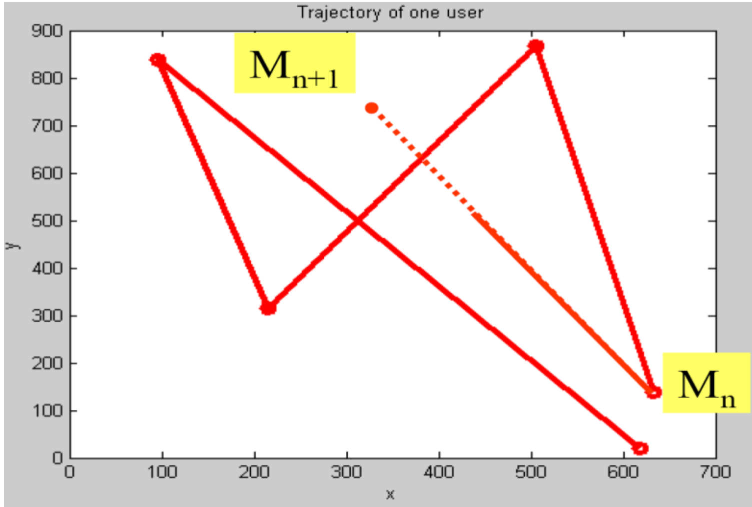
Speak the most useful mathematical language!
Espouse correct mathematical viewpoint!

Saving Private Ryan

- What is the general framework of Palm in practical terms?
- Peeling off all formalities, it is analogous to the following examples.



- Glancing at a queuing network **for each arrival event**



- Glimpse into a random way point mobility model **at waypoints**

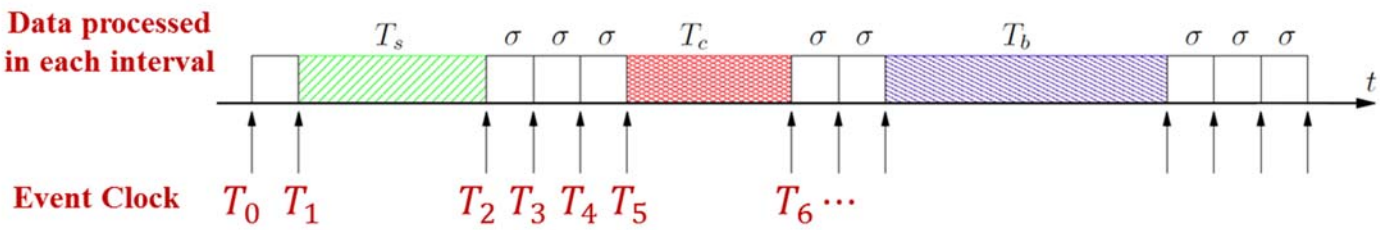


Fig. 1. The channel view of a node

- Peeping into IEEE 802.11 backoff procedure **at the beginning of time-slots**

1. Event versus Time Averages

- Consider a simulation, state of which is $\boldsymbol{S_t}$ (Jumping Process)
- Assume simulation has a **stationary** regime

- *Event clock*: times $\boldsymbol{T_n}$ at which some specific changes of state occur
  - Ex: arrival of job
  - Ex. queue becomes empty
- *Event average* statistic

$$\bar{Q}^0 := \frac{1}{N+1} \sum_{n=0}^{N} Q(T_n^-)$$

Two interpretations:
i) $Q(\cdot)$ just before T_n
ii) $\lim_{t \uparrow T_n} Q(\cdot)$

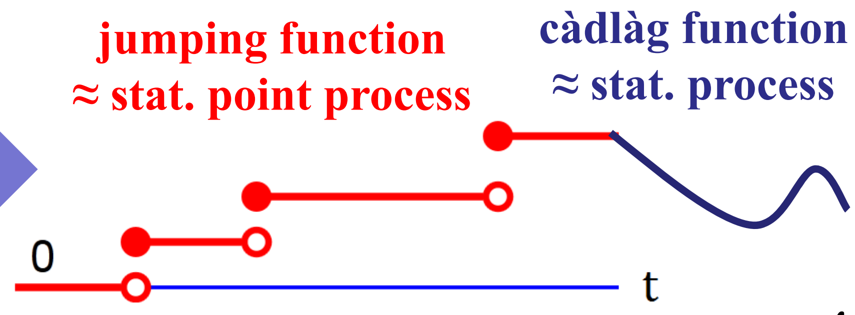
e.g., average queue length

- *Time average* statistic

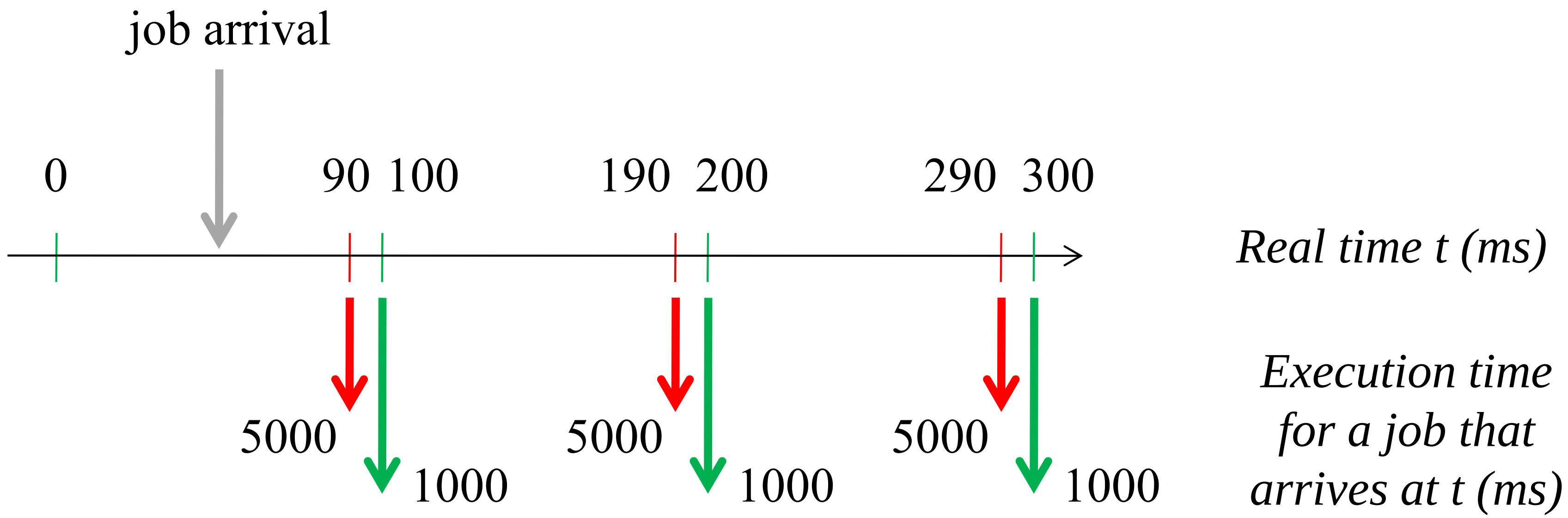
$$\bar{Q} := \frac{1}{T_N - T_0} \int_{T_0}^{T_N} Q(s)ds$$

càdlàg function

*"**c**ontinue **à** **d**roite, **l**imite **à** **g**auche" in French,*
"right continuous with left limits" in English
Property: left limit $Q(t^-) = \lim_{s \uparrow t} Q(s)$ **exists**,
right limit $Q(t^+) = \lim_{s \downarrow t} Q(s)$ **exists** and **equals to** $Q(t)$.
Skorokhod space: a collection of càdlàg functions



Example: Gatekeeper; Average execution time



Viewpoint 1: System Designer	Viewpoint 2: Customer
Two processes, with execution times 5000 and 1000	Inspector arrives at a random time
$W_s = \frac{5000 + 1000}{2} = 3000$	red processor is used with probability $\frac{90}{100}$
It looks like an improper method for this example!	$W_c = \frac{90}{100} \times 5000 + \frac{10}{100} \times 1000 = 4600$

This is a bit artificially forged example but provides some basic insights.

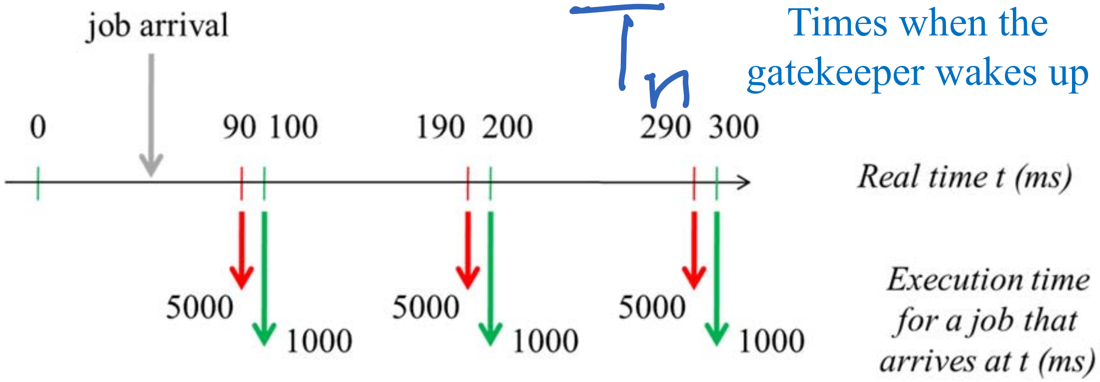
Sampling Bias

- W_s and W_c are different.
- A metric definition should mention the sampling method (*viewpoint*).
- Different sampling methods may provide different values:
  - This is the ***sampling bias***.

- *Palm Calculus* is a set of formulae for relating different viewpoints

- The relation can often be obtained by means of the *Large Time Heuristic*

Large Time Heuristic Explained on an Example



Times when the gatekeeper wakes up T_n

X_n execution time for a job processed at time T_n

- We want to relate W_s and W_c
  We apply the large time heuristic

1. How do we evaluate these metrics in a simulation ?

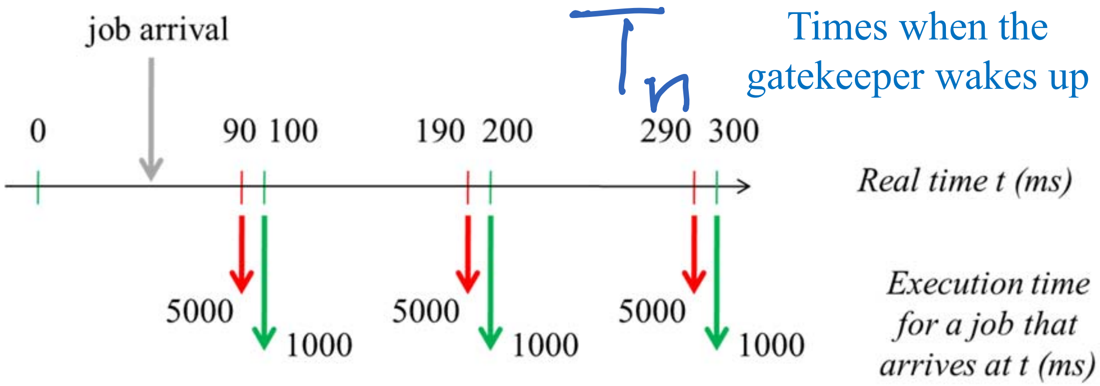
System Designer
$$W_s = \frac{1}{N} \sum_{n=1 \dots N} X_n = \bar{X}$$

Customer
$$W_c = \frac{1}{T} \int_0^T X_{N^+(t)} dt$$

$N^+(t)$: the index of the next event clock tick after t.

where $N^+(t)$ = index of next **green** or **red** arrow after t (as well as at t)

Large Time Heuristic Explained on an Example



X_n execution time for a job processed at time T_n

2. Break one integral into pieces that match the T_n's:

$$W_s = \frac{1}{N} \sum_{n=1 \ldots N} X_n = \bar{X}$$

$$W_c = \frac{1}{T} \int_0^T X_{N^+(t)} dt$$

$N^+(t)$: the <u>index</u> of the next event clock tick after t.

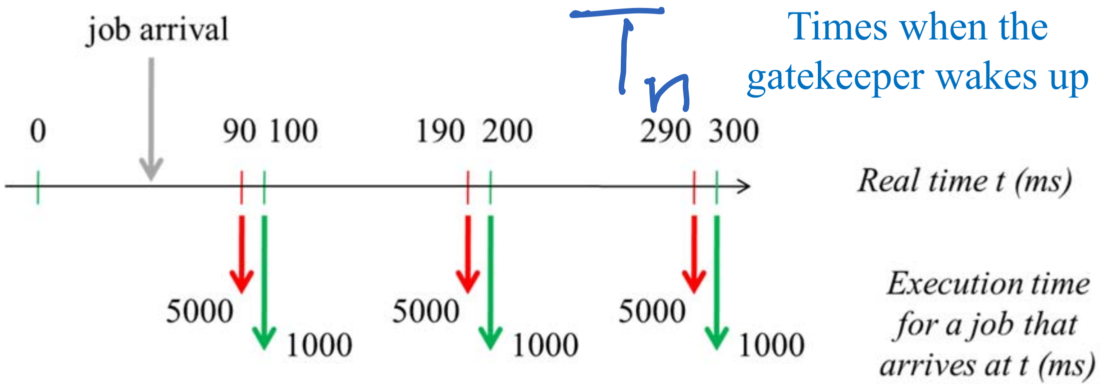
$$W_c = \frac{1}{T}\left(\int_0^{T_1} X_{N^+(t)} dt + \int_{T_1}^{T_2} X_{N^+(t)} dt + \cdots + \int_{T_{N-1}}^{T_N} X_{N^+(t)} dt \right)$$

$$= \frac{1}{T}\left(\int_0^{T_1} X_1 dt + \int_{T_1}^{T_2} X_2 dt + \cdots + \int_{T_{N-1}}^{T_N} X_N dt \right)$$

$$= \frac{1}{T}\left(T_1 X_1 + (T_2 - T_1) X_2 + \cdots + (T_N - T_{N-1}) X_N \right)$$

$$= \frac{1}{T}\left(S_1 X_1 + S_2 X_2 + \cdots + S_N X_N \right)$$

Large Time Heuristic Explained on an Example



X_n execution time for a job processed at time T_n

3. Compare (and try to **assimilate** it to W_s)

$$W_c = \frac{1}{T}(S_1X_1 + S_2X_2 + \cdots + S_NX_N)$$

$$= \frac{N}{T} \times \frac{1}{N}(S_1X_1 + S_2X_2 + \cdots + S_NX_N)$$

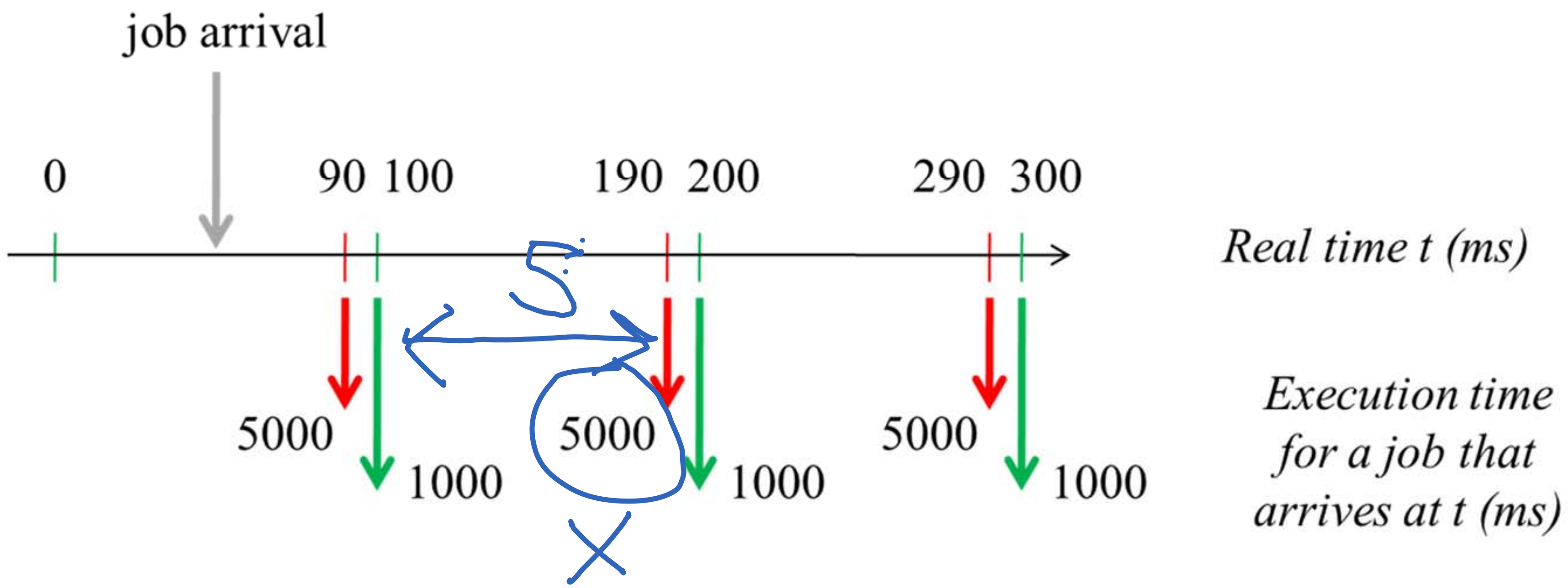
$$= \lambda \times (\mathrm{cov}(S,X) + \bar{S}\,\bar{X}) = \lambda \times \left(\mathrm{cov}(S,X) + \frac{1}{\lambda}\,\bar{X}\right)$$

$$W_c = \lambda\,\mathrm{cov}(S,X) + W_s$$

$N^+(t)$: the index of the next event clock tick after t.

Note: $\lambda = \frac{N}{T} = \frac{1}{\bar{S}}$

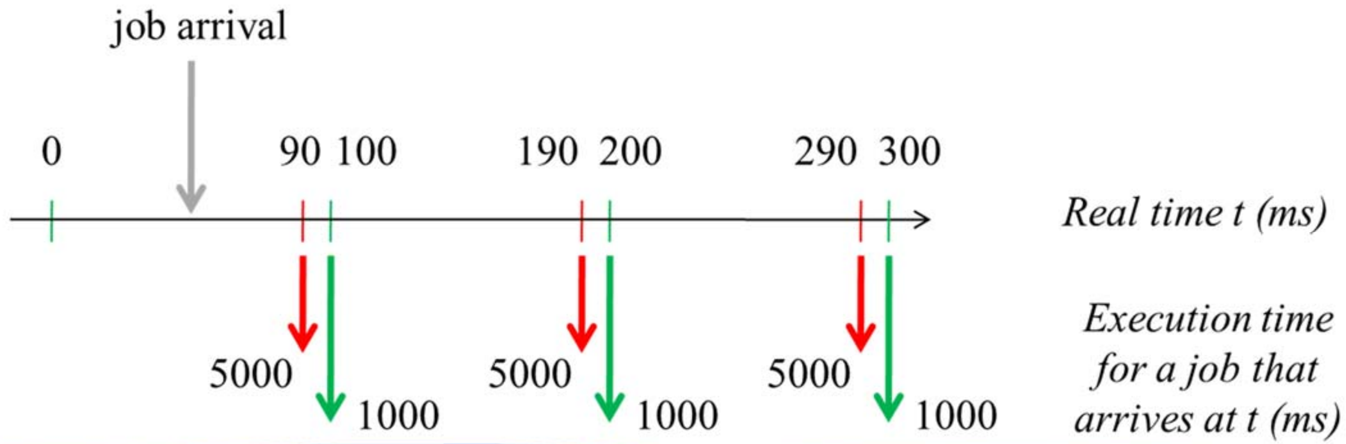
This is Palm Calculus!



$$W_c = \lambda \operatorname{cov}(S, X) + W_s$$

$$\therefore W_c \geq W_s$$

Dependency between inter-arrival time and execution time



Viewpoint 1: System Designer	Viewpoint 2: Customer
Two processes, with execution times 5000 and 1000	Inspector arrives at a random time
$W_s = \frac{5000+1000}{2} = 3000$	red processor is used with probability $\frac{90}{100}$
	$W_c = \frac{90}{100} \times 5000 + \frac{10}{100} \times 1000 = 4600$

- S_n = 90, 10, 90, 10, 90
- X_n = 5000, 1000, 5000, 1000, 5000

- Correlation is >0
- $W_c > W_s$

- **When do the two viewpoints coincide?** When either S_n or X_n is identical.

The Large Time Heuristic

1. formulate each performance metric as a long run ratio, as you would do if you would be evaluating the metric in a discrete event simulation;
2. take the formula for the time average viewpoint and break it down into pieces, where each piece corresponds to a time interval between two selected events;
3. compare the two formulations.

Large Time Heuristic:
Break a time average statistic into a sum of its event average equivalent and remaining terms.

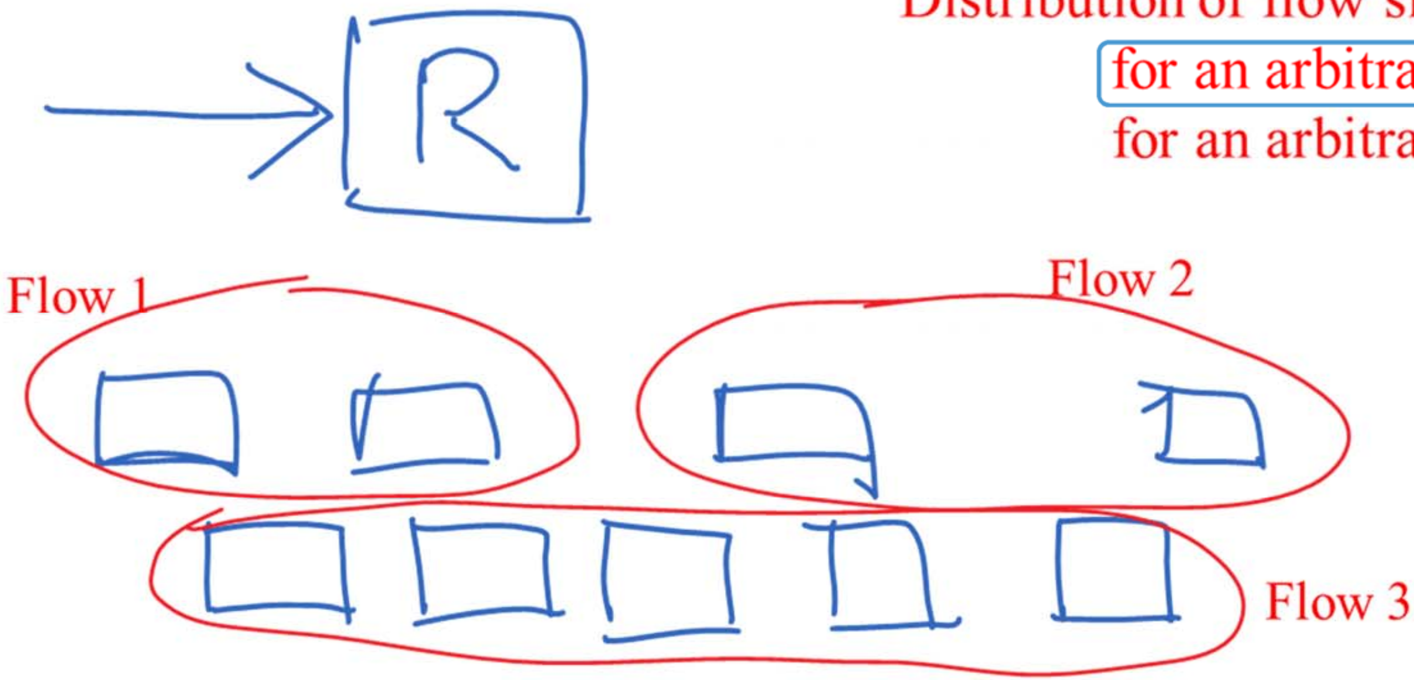
- Formally correct if simulation is stationary
  ∵ It is a corollary of Palm Inversion Formula for the case where $X(t)$ changes **only** at $t = T_n$. It can be called a **synchronized jump** case.
- It is a *robust* method, i.e., independent of assumptions on distributions (and on independence)

Other «Clocks» (in Metaphoric Sense)

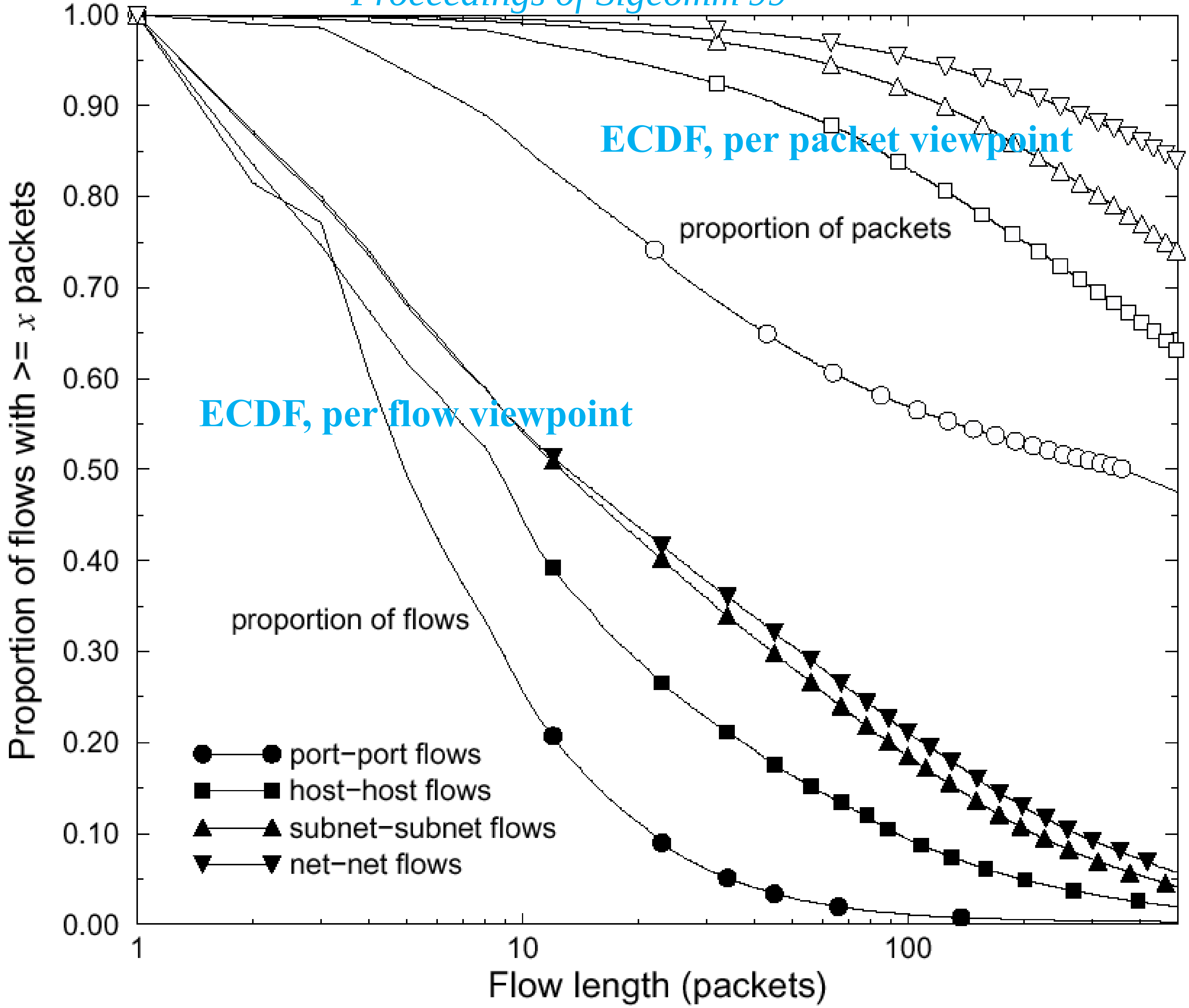
EXAMPLE 7.3: FLOW VERSUS PACKET CLOCK [96]. Packets arriving at a router are classified in "flows". We would like to plot the empirical distribution of flow sizes, counted in packets. We measure all traffic at the router for some extended period of time. Our metric of interest is the probability distribution of flow sizes. We can take a flow "clock", or viewpoint, i.e. ask: pick an arbitrary flow, what is its size ? Or we could take a packet viewpoint and ask: take an arbitrary packet, what is its size ? We have thus two possible metrics (Figure 7.3):

Our intuition tells us that this must be correct

Distribution of flow sizes
- for an arbitrary flow
- for an arbitrary packet



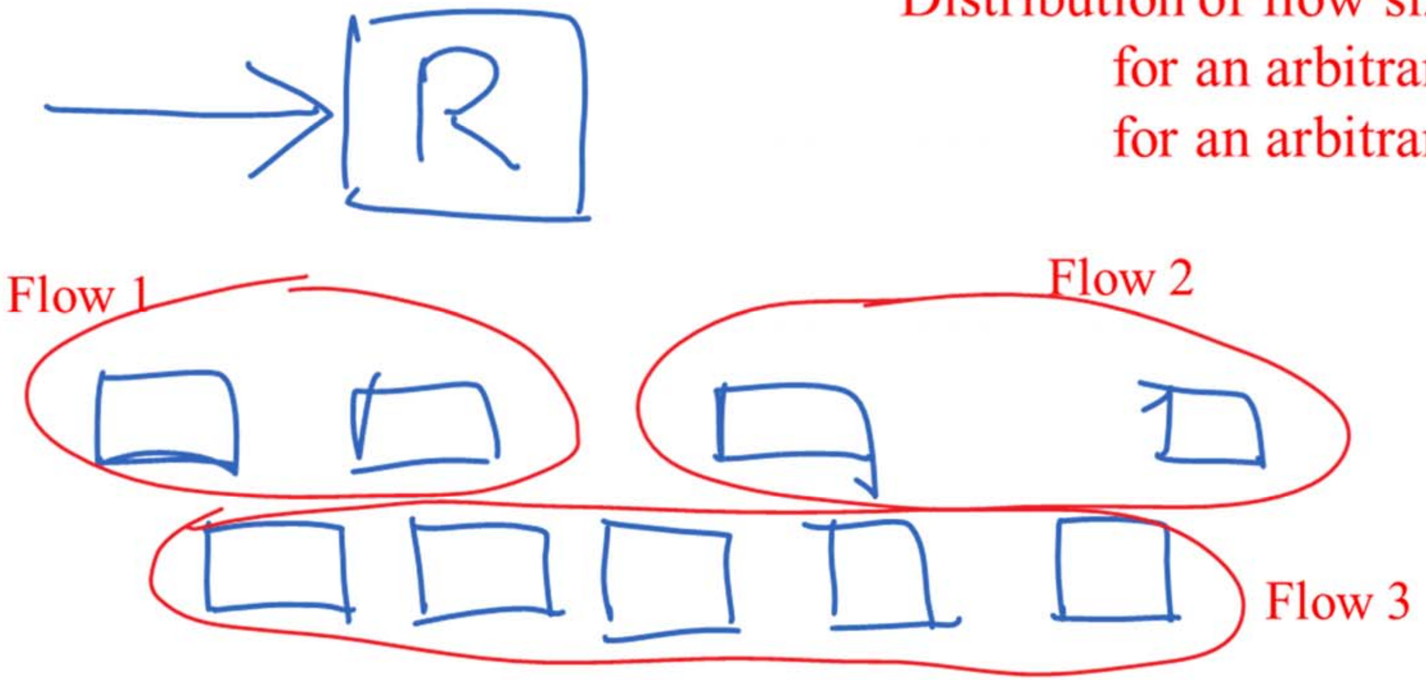
Load Sensitive Routing of Long-Lived IP Flows
Anees Shaikh, Jennifer Rexford and Kang G. Shin
Proceedings of Sigcomm'99



Distribution of flow sizes

for an arbitrary flow

for an arbitrary packet



Per flow $f_F(s) = 1/N \times$ number of flows with length s, where N is the number of flows in the dataset;

Per packet $f_P(s) = 1/P \times$ number of packets that belong to a flow of length s, where P is the number of packets in the dataset;

Mean flow size:

per flow S_F ← **Correct estimate**

per packet S_P

Large «Time» Heuristic

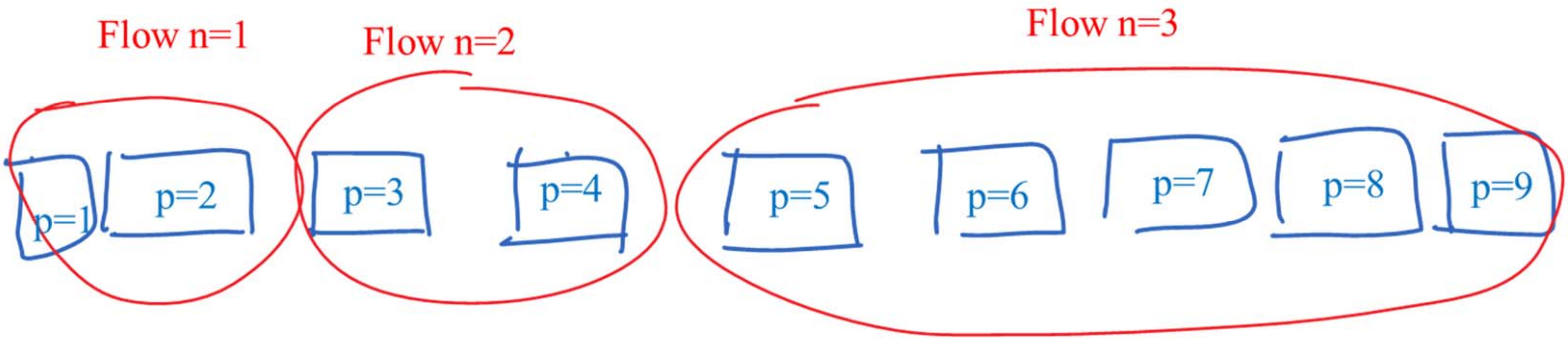
1. How do we evaluate these metrics in a simulation ?

   per flow $S_F = \frac{1}{N} \sum_n S_n$

   per packet $S_P = \frac{1}{P} \sum_p S_{F(p)}$

   where $\boldsymbol{F(p) = n}$ **when packet** $\boldsymbol{p}$ **belongs to flow** $\boldsymbol{n}$

2. **Put the packets side by side, sorted by flow**

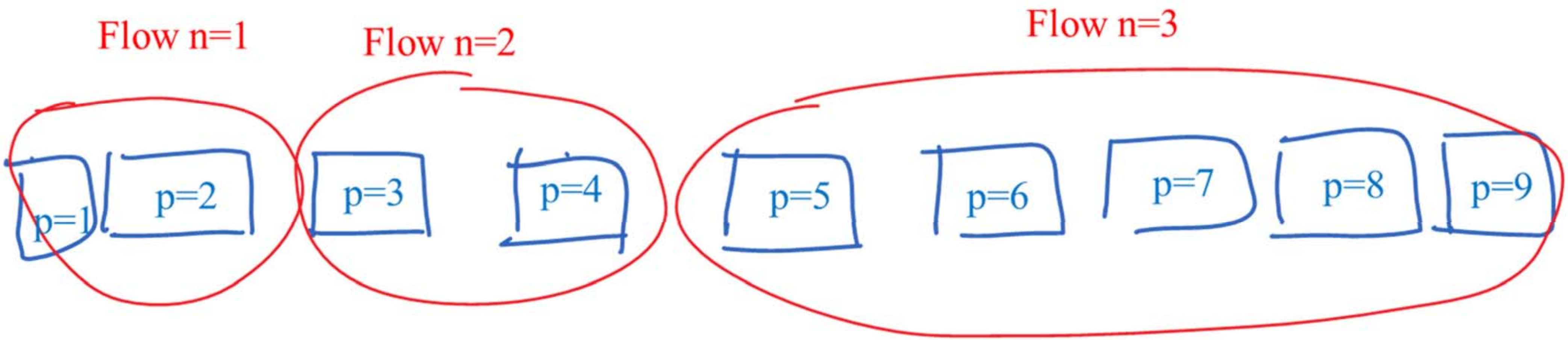


Size of Flow 3 is counted five times.

$$S_P = \frac{1}{P}(S_1 + S_1 + S_2 + S_2 + \boldsymbol{S_3 + S_3 + S_3 + S_3 + S_3} + \cdots)$$

$$= \frac{1}{P}(S_1 \times S_1 + S_2 \times S_2 + S_3 \times S_3 + \cdots) = \frac{1}{P}\sum_n S_n^2$$

Large «Time» Heuristic



3. Compare

$$S_P = \frac{1}{P}\sum_n S_n^2$$

$$S_F = \frac{1}{N}\sum_n S_n = \frac{1}{N}P$$

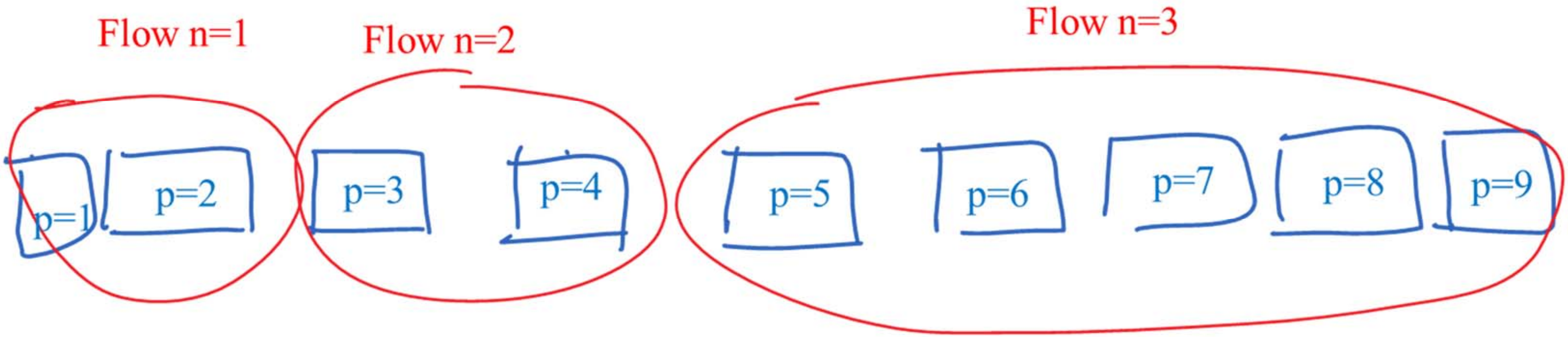
$$S_P = \frac{N}{P} \times \frac{1}{N}\sum_n S_n^2 = \frac{1}{S_F} \times \frac{1}{N}\sum_n S_n^2 = \frac{1}{S_F} \times \left(\left(\frac{1}{N}\sum_n S_n\right)^2 + var_F(S)\right) = \frac{1}{S_F} \times \left(S_F^2 + var_F(S)\right)$$

$$\boldsymbol{S_P = S_F + \frac{1}{S_F} var_F(S)}$$

Arbitrary sampling (packet average) leads to overestimation of the flow size.

Large «Time» Heuristic for PDFs of flow sizes

- Put the packets side by side, sorted by flow



1. How do we evaluate these metrics in a simulation ?

$$f_F(s) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\{S_n = s\}} \quad (7.8)$$

$$f_P(s) = \frac{1}{P}\sum_{p=1}^{P} \mathbf{1}_{\{S_{F(p)} = s\}} \quad (7.9)$$

where S_n be the size in bytes of flow n, for $n = 1, \dots N$, and $F(p)$ is the index of the flow that packet number p belongs to.

1. For s spanning the set of observed flow sizes:

S_n : size of flow n

$$f_F(s) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\{S_n=s\}} \tag{7.8}$$

$$f_P(s) = \frac{1}{P}\sum_{p=1}^{P} \mathbf{1}_{\{S_{F(p)}=s\}} \tag{7.9}$$

where S_n be the size in bytes of flow n, for $n = 1, \ldots N$, and $F(p)$ is the index of the flow that packet number p belongs to.

2. We can break the sum in Eq.(7.9) into pieces that correspond to ticks of the flow clock:

Move to simplify index condition

$$f_P(s) = \frac{1}{P}\sum_{n=1}^{N}\sum_{p:F(p)=n} \mathbf{1}_{\{S_n=s\}} = \frac{1}{P}\sum_{n=1}^{N}\sum_{p=1}^{P} \mathbf{1}_{\{F(p)=n\}}\mathbf{1}_{\{S_n=s\}}$$

A bit hard to hit upon this idea

$$= \frac{1}{P}\sum_{n=1}^{N}\mathbf{1}_{\{S_n=s\}}\sum_{p=1}^{P}\mathbf{1}_{\{F(p)=n\}} = \frac{1}{P}\sum_{n=1}^{N}\mathbf{1}_{\{S_n=s\}}s = \frac{s}{P}\sum_{n=1}^{N}\mathbf{1}_{\{S_n=s\}} \tag{7.10}$$

There are s packets in Flow n.

3. Compare Eqs.(7.8) and (7.10) and obtain that for all flow size s:

$$f_P(s) = \eta s f_F(s) \tag{7.11}$$

where η is a normalizing constant ($\eta = N/P$).

Packet clock based samples lead to a heavier tail.

Cyclist's Paradox

- On a round trip tour, there is more uphills than downhills

EXAMPLE 7.4: KILOMETER VERSUS TIME CLOCK: CYCLIST'S PARADOX. A cyclist rides swiss mountains; his speed is 10 km/h uphill and 50 km/h downhill. A journey is made of 50% uphill slopes and 50% downhill slopes. At the end of the journey, the cyclist is disappointed to read on his speedometer an average speed of only 16.7 km/h, as he was expecting an average of $\frac{10+50}{2} = 30$ km/h.

Or running on treadmills?

The km clock vs. the standard clock

- v_ℓ = speed for the ℓ^{th} kilometer (the trip consists of equal-sized pieces, 1,…,L)

$$S_{\text{kilometer}} = \frac{1}{L}\sum_{\ell} v_\ell = \textbf{mean of } \boldsymbol{v_\ell}$$

$$S_{\text{time}} = \frac{L}{T} = \frac{L}{\sum_\ell \frac{1}{v_\ell}} = \textbf{harmonic mean of } \boldsymbol{v_\ell} < \textbf{mean of } \boldsymbol{v_\ell}$$

Using the same method as in Example 7.3, one obtains

standard clock $$f_t(v) = \eta \frac{1}{v} f_k(v) \qquad (7.14)$$ **kilometer clock**

where $f_t(v)$ [resp. $f_k(v)$] is the PDF of the speed, sampled with the standard clock [resp. km clock] and η is a normalizing constant; f_t puts more mass on the small values of the speed v, this is another explanation to the cyclist's paradox.

2. Palm Calculus : Framework

- A stationary process (simulation) with state S_t
- Some quantity X_t measured at time t. Assume that

(S_t,X_t) is stationary

i.e., S_t is in a stationary regime and X_t **may depend** on the past, present and future state of the simulation S_t in a way that is (probabilistically) invariant by shift of time origin.

- Examples : **S_t can be any state of a simulation**
  - S_t = current position of mobile, speed, and next waypoint
  - *X_t jointly stationary with S_t*:
    - X_t = current speed at time t; X_t = remaining time until next waypoint
  - *X_t not jointly stationary with S_t*:
    - X_t = **absolute** time at which last waypoint occurred

Stationary Point Process

- Consider some selected transitions of the simulation, occurring at times T_n.
  - Example: T_n = time when n^{th} trip ends in random waypoint model

Formally, a stationary point process in our setting is associated with a subset $\mathcal{F}_0$ of the set of all possible state transitions of the simulation. It is made of all time instants t at which the simulation does a transition in $\mathcal{F}_0$, i.e. such that $(S(t^-), S(t^+)) \in \mathcal{F}_0$.

state transition

- T_n is a called a *stationary* ***point*** *process* associated to S_t
  - T_n grows over time, so think of it as a set of "***points***" generated by T_n
  - Stationary because S_t is stationary
  - Jointly stationary with S_t

We denote the time instant of the point process such that
$$\dots < T_{-2} < T_{-1} < T_0 \leq 0 < T_1 < T_2 \dots$$

- Time 0 is an *arbitrary* point in time, rather than the beginning of a simulation.
  - ∵ The simulation has run **for so long a time at $t = 0$** that it is now stationary!

Palm Expectation

- Assume: X_t, S_t are jointly stationary, T_n is a stationary point process **associated** with S_t.
- ***Definition*** : the **Palm Expectation** is

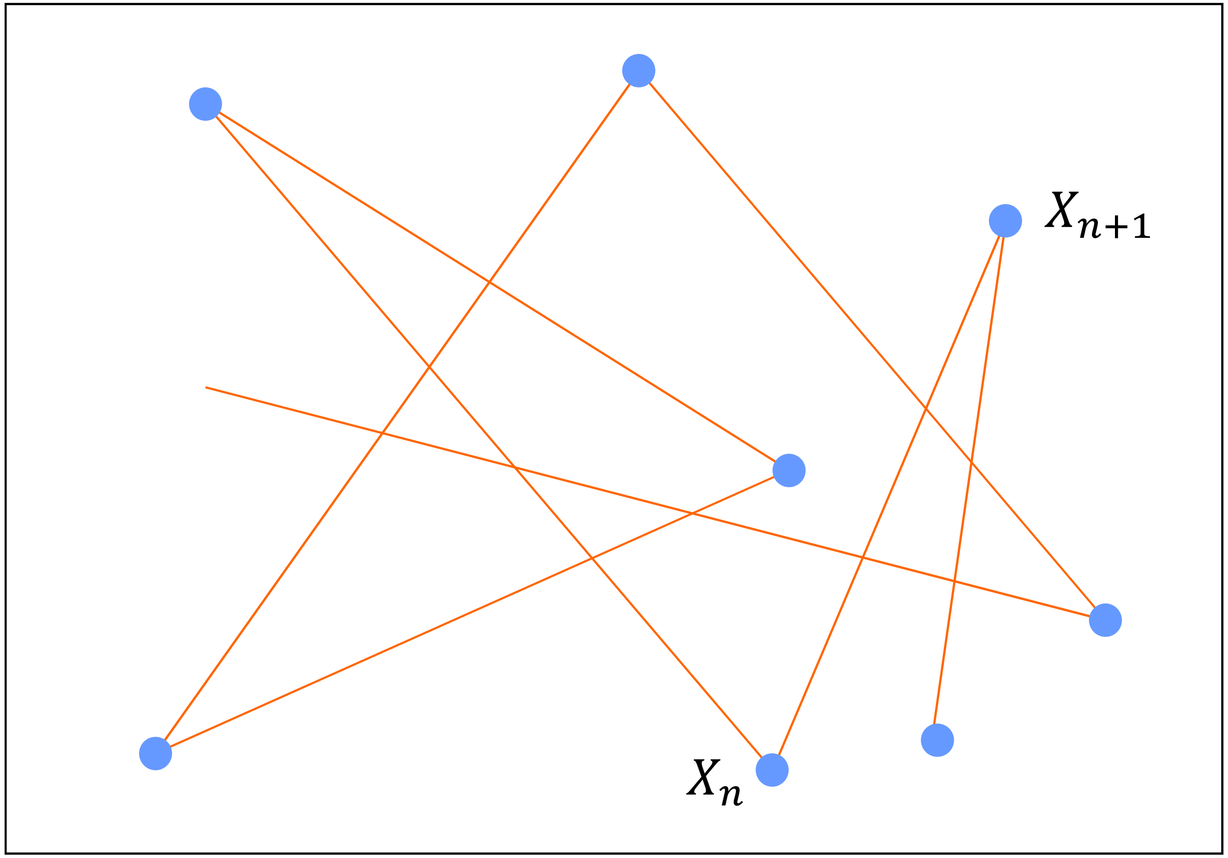
$$\mathrm{E}^t(X_t) = \mathrm{E}(X_t \mid \text{a selected transition occurred at time } t)$$

- By stationarity:

$$\mathrm{E}^t(X_t) = \mathrm{E}^0(X_0) \qquad \text{for all } t$$

- Example:
  - T_n = time when n^{th} trip ends, X_t = instant speed at time t
  - $\mathrm{E}^t(X_t) = \mathrm{E}^0(X_0)$ = average speed observed at a waypoint

- $E(X_t) = E(X_0)$ expresses the time average viewpoint.
- $E^t(X_t) = E^0(X_0)$ expresses the event average viewpoint.
- Example for random waypoint:
  - T_n = time when n^{th} trip ends, X_t = instant speed at time t
  - $E^t(X_t) = E^0(X_0)$ = average speed observed **at trip ends**
  - $E(X_t) = E(X_0)$ = average speed observed **at an arbitrary point in time**



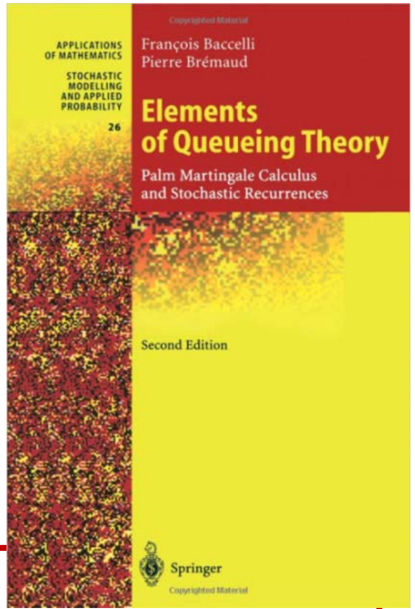
A Palm expectation is always associated with a stationary point process T_n.

Formal Definition

- In **discrete time**, we have a definition based on an elementary conditional probability.

$$\mathbb{E}^t(Y) = \mathbb{E}(Y|N(t)=1) = \frac{\mathbb{E}(YN(t))}{\mathbb{E}(N(t))} = \frac{\mathbb{E}(YN(t))}{\mathbb{P}(N(t)=1)}$$

- In **continuous time**, the definition is **appallingly** sophisticated
  - *Radon-Nykodim* derivative – see textbook for details.
  - Also see [**3, 4**] for a formal treatment.



- Palm probability is defined similarly.

The Palm *probability* is defined similarly, namely

$$\mathbb{P}^0(X(0)\in W) = \mathbb{P}(X(0)\in W|\text{ a point of the process } T_n \text{ occurs at time } 0)$$

for any measurable subset W of the set of values of $X(t)$. In particular, we can write $\mathbb{P}^0(T_0 = 0) = 1$.

We denote the time instant of the point process such that
$\ldots < T_{-2} < T_{-1} < T_0 \leq 0 < T_1 < T_2 \ldots$

Ergodic Interpretation

- Assume simulation is stationary + ergodic, i.e., sample path averages converge to expectations; then we can estimate time and event averages by:

$$\mathbb{E}(\phi(X(t))) \approx \frac{1}{T}\sum_{t=1}^{T}\phi(X(t))$$

$$\mathbb{E}^0\left(\phi(X(0))\right) \approx \frac{1}{N}\sum_{n=1}^{N}\phi\left(X\left(T_n\right)\right)$$

Lack of ergodicity implies these formulae do not hold.

e.g., disconnected Markov chain

- In terms of probabilities:
  - Stationary probability:
    $\mathbb{P}(X_t \in W) \approx$ fraction of time that X_t is in some set W
  - Palm probability:
    $\mathbb{P}^t(X_t \in W) \approx$ fraction of selected transitions at which X_t is in W

Intensity of a Stationary Point Process

- Intensity of selected transitions:
  - **λ := expected number of transitions per time unit**

INTENSITY The *intensity* λ of the point process is defined as the expected number of points per time unit. We have assumed that there cannot be two points at the same instant. In discrete or continuous time, the intensity λ is defined as the unique number such that the number $N(t, t+\tau)$ of points during any interval $[t, t+\tau]$ satisfies [4]:

General definition $$\mathbb{E}(N(t, t+\tau)) = \lambda\tau \quad (7.16)$$

In discrete time, λ is also simply equal to the probability that there is a point at an arbitrary time:

Discrete-time version $$\lambda = \mathbb{P}(T_0 = 0) = \mathbb{P}(N(0) = 1) = \mathbb{P}(N(t) = 1) \quad (7.17)$$

where the latter is valid for any t, by stationarity.

One can think of λ as the (average) rate of the event clock.

Wordy yet consistent and intuitive definition

Which is zero for continuous-time case

Two Palm Calculus Formulae

- **Intensity Formula**:

We denote the time instant of the point process such that
$\ldots < T_{-2} < T_{-1} < T_0 \le 0 < T_1 < T_2 \ldots$

$$\frac{1}{\lambda} = \mathbb{E}^0(T_1 - T_0) = \mathbb{E}^0(T_1)$$

$T_0 = 0$

where by convention $T_0 \le 0 < T_1$

- **(Palm) Inversion Formula** (a.k.a. Ryll-Nardzewski and Slivnyak's formula)

Any time average!
1. Avg. Throughput
2. Avg. Customer

$$\mathbb{E}(X_t) = \mathbb{E}(X_0) = \lambda \mathbb{E}^0\left(\int_0^{T_1} X_s ds\right)$$

Time average of X_t (also denoted as $X(t)$)

average of X_t between two events, $T_0 = 0$ and T_1

- The proofs are simple in discrete time – see textbook

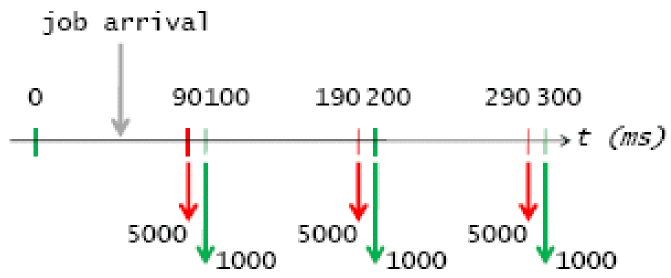
The only assumption is *stationarity*, dispensing with independence or Poisson assumptions. Once again, do not forget that T_n is *another* point process, only jointly stationary with X_t.

EXAMPLE 7.5: GATEKEEPER, CONTINUED. Assume we model the gatekeeper example as a discrete event simulation, and consider as point process the waking ups of the gatekeeper. Let $X(t)$ be the execution time of a hypothetical job that would arrive at time t. The average job execution time, sampled with the standard clock (customer viewpoint) is

$$W_c = \mathbb{E}(X(t)) = \mathbb{E}(X(0))$$

whereas the average execution time, sampled with the event clock (system designer viewpoint), is

$$W_s = \mathbb{E}^t(X(t)) = \mathbb{E}^0(X(0))$$



The inversion formula gives

Time average of X_t $$W_c = \lambda \mathbb{E}^0 \left(\int_0^{T_1} X(t)dt \right) = \lambda \mathbb{E}^0 \left(X(0) T_1 \right)$$

average of X_t between $T_0 = 0$ and T_1

(recall that T_0 =0 under the Palm probability and $X(0)$ is the execution time for a job that arrives just after time 0). Let C be the cross-covariance between sleep time and execution time:

$$C := \mathbb{E}^0(T_1 X(0)) - \mathbb{E}^0(T_1)\mathbb{E}^0(X(0))$$

then

$$W_c = \lambda \left[C + \mathbb{E}^0(X(0))\mathbb{E}^0(T_1) \right]$$

By the inversion formula $\lambda = \frac{1}{\mathbb{E}^0(T_1)}$ thus

$$W_c = W_s + \lambda C$$

which is the formula we had derived using the heuristic in Section 7.1.

3. Other Palm Calculus Formulae

THEOREM 7.3.1. *Let $X(t) = T^+(t) - t$ (time until next point, also called residual time), $Y(t) = t - T^-(t)$ (time since last point), $Z(t) = T^+(t) - T^-(t)$ (duration of current interval). For any t, the distributions of $X(t)$ and $Y(t)$ are equal, with PDF:*

$$f_X(s) = f_Y(s) = \lambda \mathbb{P}^0(T_1 > s) = \lambda \int_s^{+\infty} f_T^0(u)du \tag{7.28}$$

where f_T^0 is the Palm PDF of $T_1 - T_0$ (PDF of inter-arrival times). The PDF of $Z(t)$ is

$$f_Z(s) = \lambda s f_T^0(s) \tag{7.29}$$

In particular, it follows that

Rather than $\frac{1}{2\lambda}$ (in line with our intuition)

Recall

$$\frac{1}{\lambda} = \mathbb{E}^0(T_1)$$

$$\mathbb{E}(X(t)) = \mathbb{E}(Y(t)) = \frac{\lambda}{2}\mathbb{E}^0(T_1^2) \text{ in continuous time} \tag{7.30}$$

$$\mathbb{E}(X(t)) = \mathbb{E}(Y(t)) = \frac{\lambda}{2}\mathbb{E}^0(T_1(T_1+1)) \text{ in discrete time} \tag{7.31}$$

$$\mathbb{E}(Z(t)) = \lambda \mathbb{E}^0(T_1^2) \tag{7.32}$$

Z (current interval) is heavier than $T = T_1 - T_0$ (inter-arrival times)

The larger the interval is, the more you are likely to fall there

$$f_Z(s) = \lambda s f_T^0(s)$$

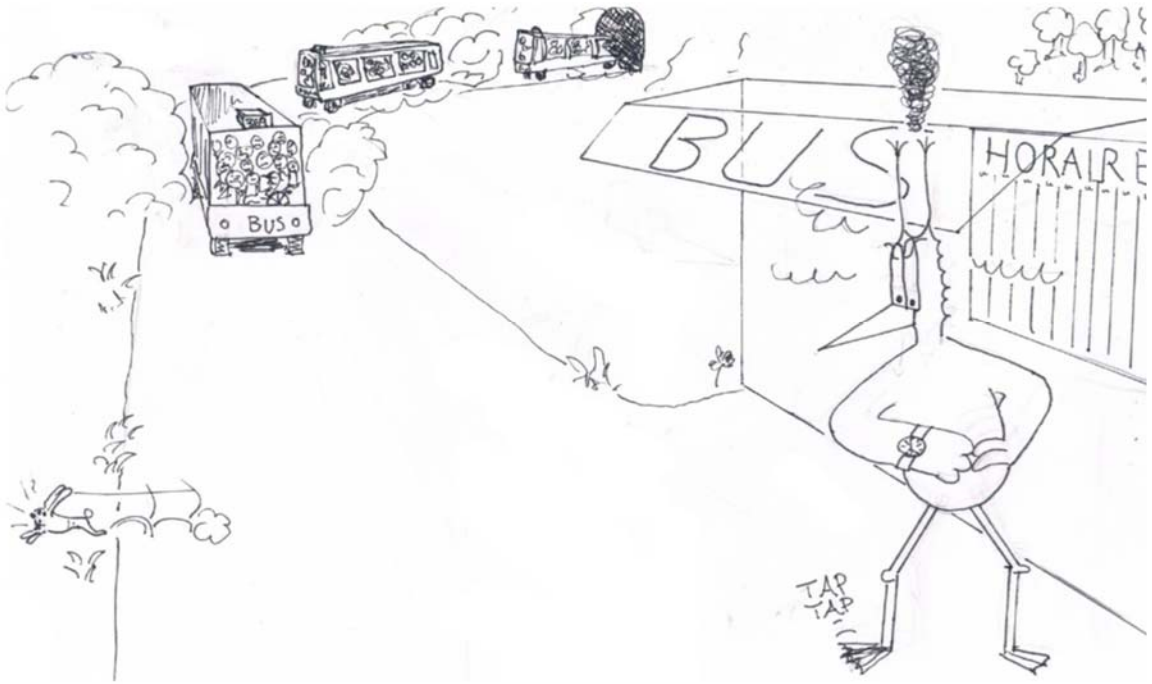
$\boldsymbol{Z(t) = T^+(t) - T^-(t)}$ **: duration of current interval**

- Density of Z, current interval, at $Z = s$ is proportional to

$$\boldsymbol{s} \times \text{Density of } \boldsymbol{T} \text{ at } \boldsymbol{T = s}$$

- In other words, the probability that you fall within an interval of size s is proportional to s times the probability density function of the intervals of the point process.

- You are likely to fall in larger intervals.



Joe's Waiting Time

$$\mathbb{E}(X(t)) = \mathbb{E}(Y(t)) = \frac{\lambda}{2}\mathbb{E}^0(T_1^2)$$

Residual time — **Elapsed time**

Recall

$$\frac{1}{\lambda} = \mathbb{E}^0(T_1)$$

- $E(X(t)) = \frac{\lambda}{2}E^0(T_1^2) = \frac{\lambda}{2}(E^0(T_1))^2 + \frac{\lambda}{2}\text{var}^0(T_1)$

  mean waiting time $= \frac{1}{2}E^0(T_1) + \frac{\lambda}{2}\text{var}^0(T_1)$

  0.5 × mean time between buses system's viewpoint

  penalty due to variability

Feller's Paradox

- At bus stop in average λ buses per hour. Inspector measures time between all bus inter-departures. Inspector estimates $\mathbb{E}^0(T_1 - T_0) = \frac{1}{\lambda}$
- Joe arrives at time t and measures $X_t =$ (time until next bus $-$ time since last bus). Joe estimates $\mathbb{E}(X_0) = \mathbb{E}(T_1 - T_0)$

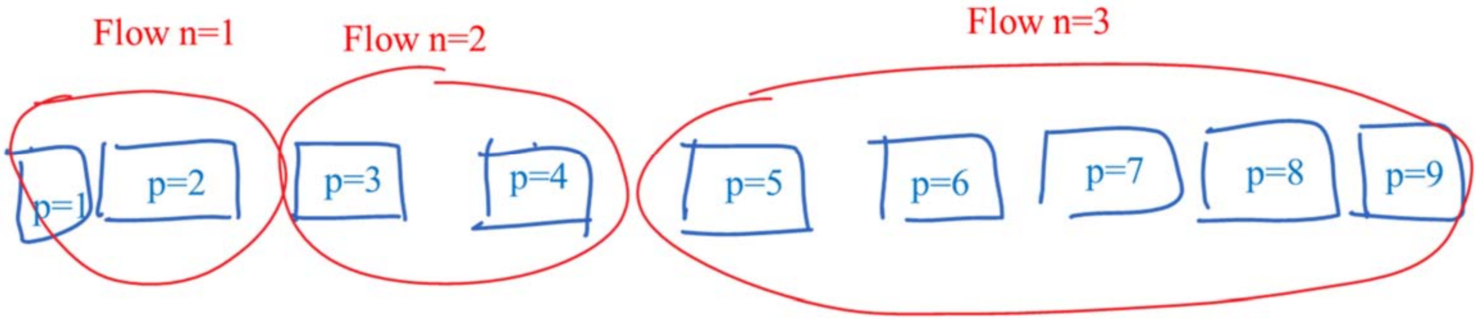
  Time average of inter-arrival time $T_1 - T_0$
- Inversion formula:

$$\mathbb{E}(T_1 - T_0) = \lambda \mathbb{E}^0\left(\int_0^{T_1} X_t dt\right) = \lambda \mathbb{E}^0(T_1^2) = \frac{1}{\lambda} + \underbrace{\lambda \text{var}^0(T_1 - T_0)}_{\text{penalty}}$$

- Joe's estimate always larger than Inspector's (Feller's Paradox)

We encountered Feller's Paradox Already

Large «Time» Heuristic



3. Compare

$$S_P = \frac{1}{P}\sum_n S_n^2$$

$$S_F = \frac{1}{N}\sum_n S_n = \frac{1}{N}P$$

$$S_P = \frac{N}{P}\times\frac{1}{N}\sum_n S_n^2 = \frac{1}{S_F}\times\frac{1}{N}\sum_n S_n^2 = \frac{1}{S_F}\times\left(\left(\frac{1}{N}\sum_n S_n\right)^2 + var_F(S)\right) = \frac{1}{S_F}\times\left(S_F^2 + var_F(S)\right)$$

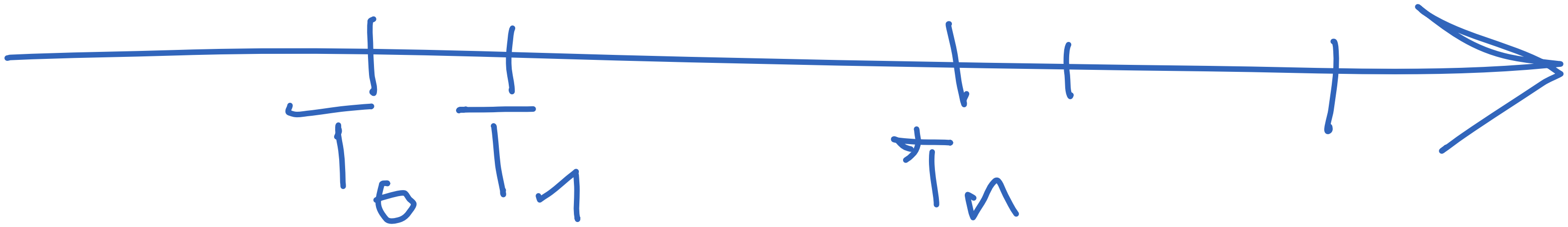
$$\boldsymbol{S_P = S_F + \frac{1}{S_F} var_F(S)}$$

Arbitrary sampling (packet average) leads to overestimation of the flow size.

16

The larger the interval is, the more you are likely to fall there → **The bigger the flow is, the more packets you are likely to sample from the flow**

For a Poisson process, what is the mean length of an interval ?



EXAMPLE 7.7: POISSON PROCESS. Assume that T_n is a Poisson process (see Section 7.6). We have $f_T^0(t) = \lambda e^{-\lambda s}$ and $\mathbb{P}^0(T_1 > s) = \mathbb{P}^0(T_1 \geq s) = e^{-\lambda s}$ thus $f_X(s) = f_Y(s) = f_T^0(s)$.

This is expected, by the memoriless property of the Poisson process: we can think that at every time slot, of duration dt, the Poisson process flips a coin and, with probability λdt, decides that there is an arrival, independent of the past. Thus, the time $X(t)$ until the next arrival is independent of whether there is an arrival or not at time t, and the Palm distribution of $X(t)$ is the same as its time average distribution. Note that this is special to the Poisson process; processes that do not have the memoriless property do not have this feature.

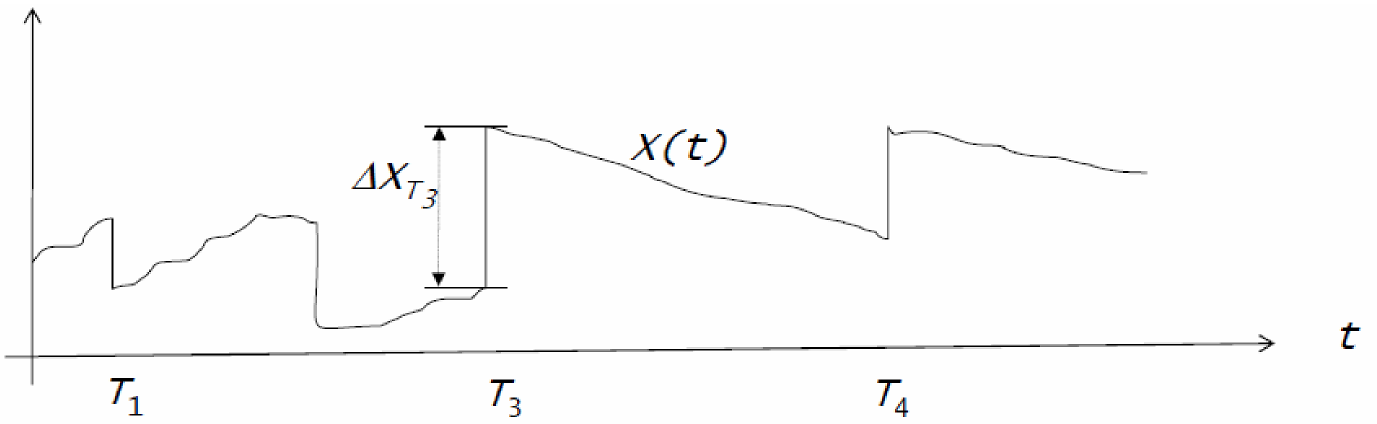
The distribution of $Z(t)$ has density

duration of current interval

$$f_T^0(s) = \lambda^2 s e^{-\lambda s}$$

i.e., it is an Erlang-2 distribution[1].

Miyazawa's Rate Conservation Law



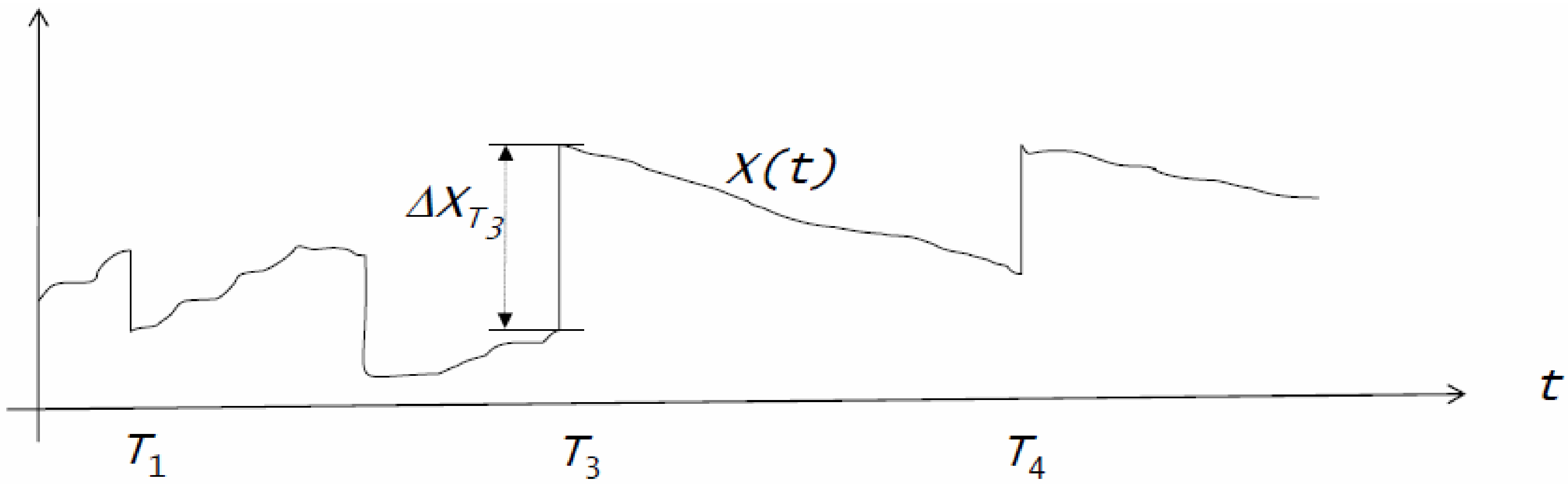
Consider a random, real valued stochastic process $X(t)$ with the following properties (Figure 7.6):

- $X(t)$ is continuous everywhere except perhaps at instants of a stationary point process T_n;
- $X(t)$ is continuous to the right; **Càdlàg function** **Jump ups and downs by point process**
- $X(t)$ has a right-handside derivative $X'(t)$ for all values of t.

Define ΔX_t by ΔX_t=0 if t is not a point of the point process T_n and $\Delta X_{T_n} = X(T_n) - X(T_n^-)$, i.e. ΔX_t is the amplitude of the discontinuity at time t. Note that it follows that

$$X(t) = X(0) + \int_0^t X'(s)ds + \sum_{n \in \mathbb{N}} \Delta_{T_n} \mathbf{1}_{\{T_n \le t\}} \qquad (7.34)$$

continuous part **discontinuous jumps**



THEOREM 7.3.2. *(Rate Conservation Law [69]) Assume that the point process T_n and $X(t)$ are jointly stationary. If $\mathbb{E}^0 |\Delta_0| < \infty$ and $\mathbb{E}|X'(0))| < \infty$ then*

Taking differentiation & expectation of the previous equation yields:

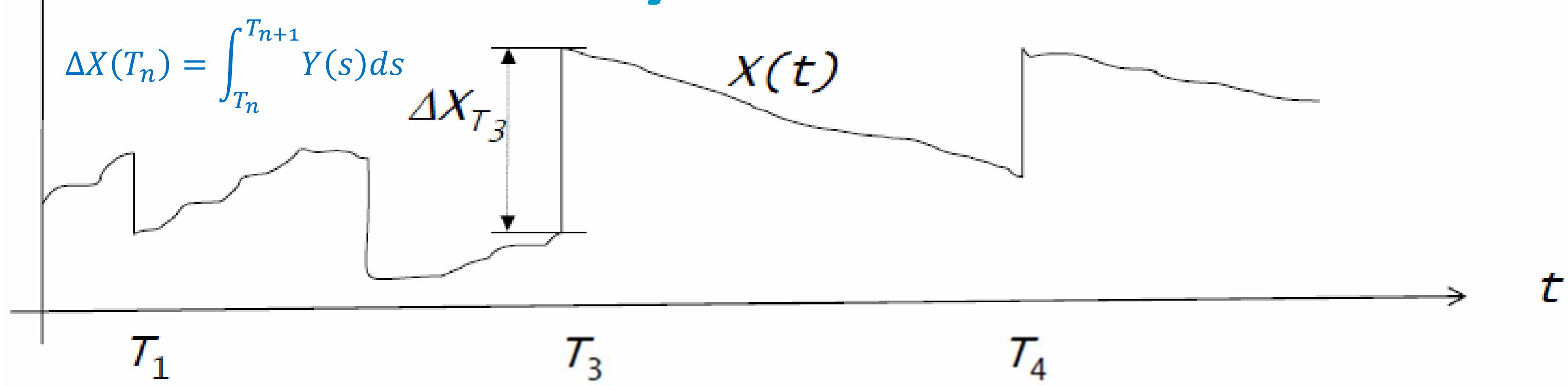
$$\mathbb{E}(X'(0)) + \lambda \mathbb{E}^0(\Delta_0) = 0$$

Palm inversion formula

where λ is the intensity of the point process T_n and E^0 is the Palm expectation.

- $\mathbb{E}(X'(0))$ (also equal to $\mathbb{E}(X'(t))$ for all t) is the average rate of increase of the process $X(t)$, excluding jumps.
- $\mathbb{E}^0(\Delta_0)$ is the expected amplitude of one arbitrary jump. Thus $\lambda \mathbb{E}^0(\Delta_0)$ is the expected rate of increase due to jumps.
- The theorem says that, if the system is stationary, the sum of all jumps cancels out, in average.

From Miyazawa to Palm



Roughly speaking, **Palm is a special case of Miyazawa**:

Let us first define:

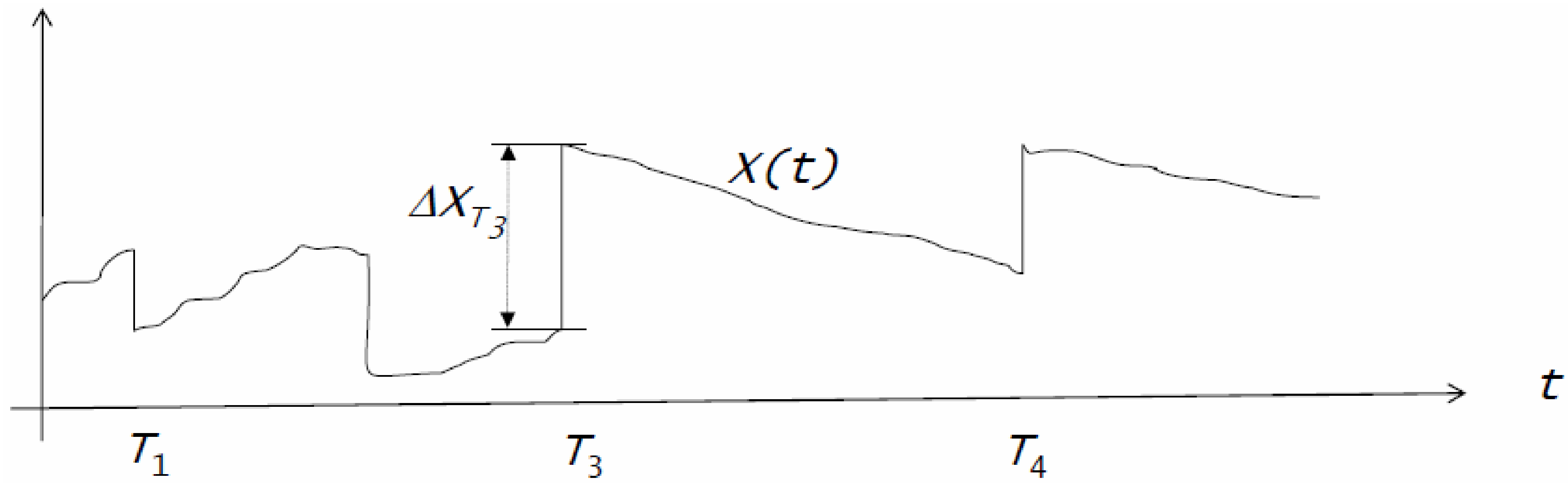
$$X(t) = \int_{t}^{T_+(t)} Y(s)ds$$

Then you have $X'(t) = -Y(t)$ and $Y_{0^-} = 0$ since $T^+(0^-) = T_0$. Hence:

$$E[X'(0)] = -E[Y(0)] = -\lambda E^0[\Delta X(0)] = -\lambda E^0\left[\int_0^{T_1} Y(s)ds\right]$$

Palm Inversion Formula

Miyazawa: Event average of jumps is counteracted by its right-hand side derivative.
Palm: Time average is equivalent to its average over one interval.



EXAMPLE 7.11: POWER CONSUMPTION PER JOB. A system serves jobs and consumes in average $\bar{P}$ watts. Assume we allocate the energy consumption to jobs, for example by measuring the current when a job is active. Let $\bar{E}$ be the total energy consumed by a job, during its lifetime, in average per job, measured in Joules. By Eq.(7.40):

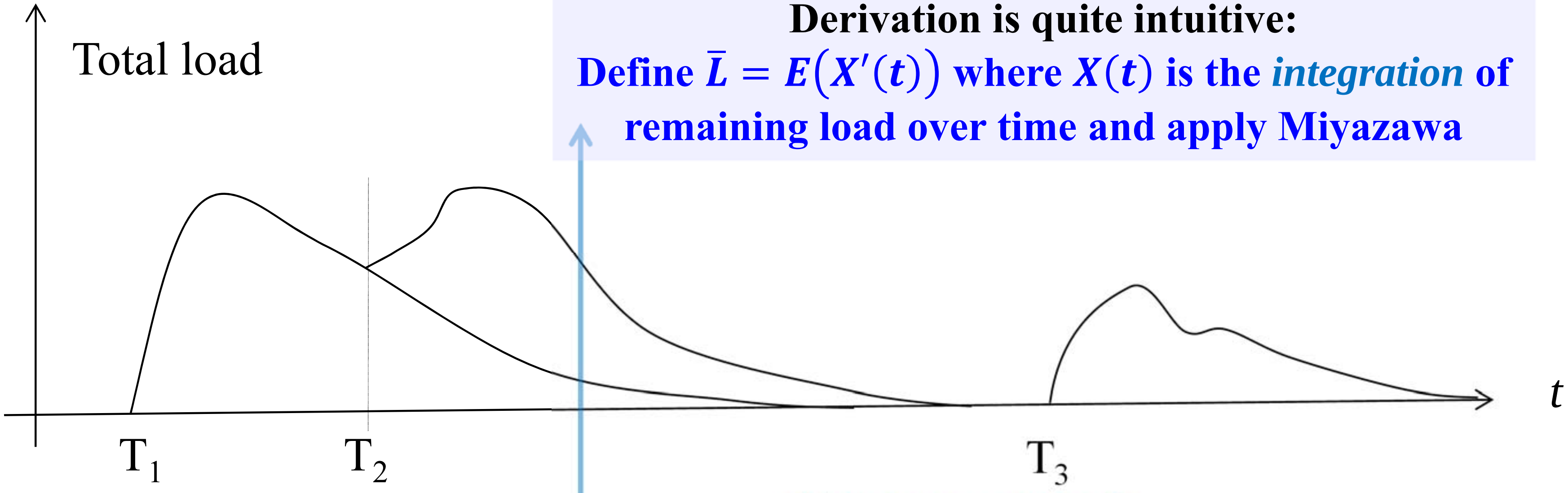
$$\bar{P} = \lambda \bar{E}$$

where λ is the number of jobs per second, served by the system.

$$\mathbb{E}\left(X'(0)\right) + \underbrace{\lambda \mathbb{E}^0\left(\Delta_0\right)}_{\text{Palm inversion formula}} = 0$$

Campbell's Formula

Derivation is quite intuitive:
Define $\bar{L} = E(X'(t))$ where $X(t)$ is the *integration* of remaining load over time and apply Miyazawa



cf. instantaneous ones so far

- Shot Noise Model: customer n adds an ***arbitrarily dispersed*** load $h(t - T_n, Z_n)$ where Z_n is some ***random*** attribute and T_n is arrival time

THEOREM 7.3.3 (Shot Noise). *The average load at an arbitrary point in time is*

$$\bar{L} = \lambda \times \mathbb{E}^0 \left(\int_0^\infty h(t, Z_0) dt \right) \tag{7.39}$$

where equality holds also if either $\bar{L}$ or the work per customer is infinite.

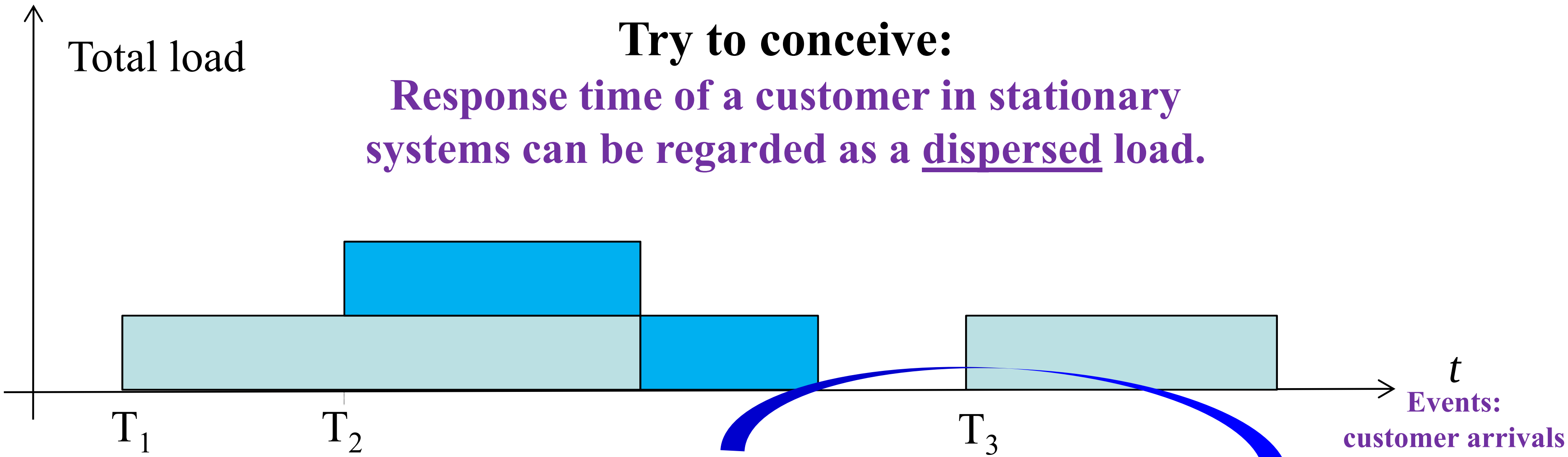
- Trivial example: Throughputs of TCP flows, $L = \lambda V$ with L = **bits per second**, V = **total bits per flow** and λ= **flows per sec**

Little's Law is merely an immediate consequence of Campbell's Formula.

Little’s Formula

Try to conceive:

Response time of a customer in stationary systems can be regarded as a <u>dispersed</u> load.



We can apply Campbell’s formula by letting $Z_n = R_n$ and $h(t, z) = \mathbf{1}_{\{0 \le t < z\}}$, i.e. the load generated by one customer is 1 as long as it is present in the system; equivalently, we can apply the rate conservation law with $X(t)$ = residual time to be spent by customers present in in the system. This gives the celebrated theorem:

Integral of which becomes ‘response time’

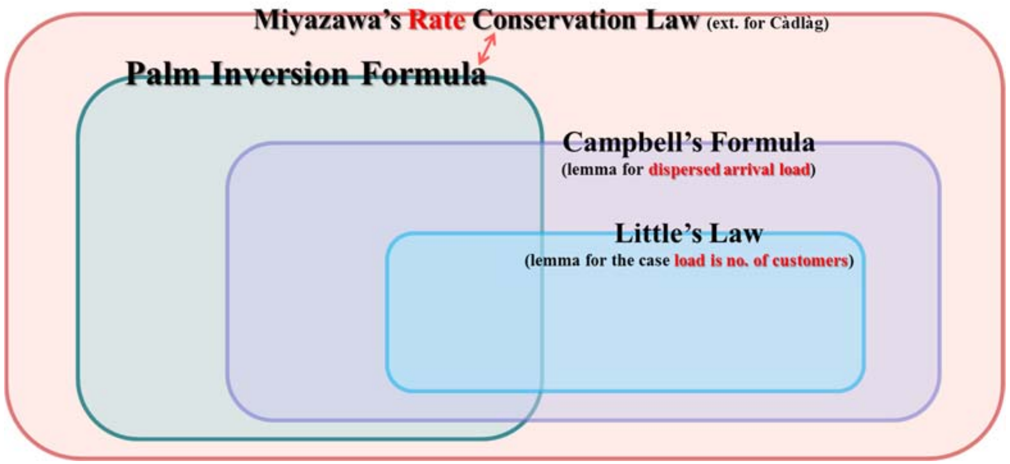
THEOREM 7.3.4 (*Little’s Formula*). *The mean number of customers in the system at time t,* $\bar{N} := \mathbb{E}(N(t))$*, is independent of* t *and satisfies*

$$\bar{N} = \lambda \bar{R}$$

where λ *is the arrival rate and* $\bar{R}$ *the average response time, experienced by an arbitrary customer.*

Apply Campbell’s Formula to the time average $N(t) = \sum_n \mathbf{1}_{\{T_n \le t < T_n + R_n\}}$.

High-level Recap



All these theorems essentially implies *conservation*:
Event average of jumps must be counteracted by its time average derivative.

- **Dispersed load** in Campbell's Formula
  - Coined term for a generalized or **reinterpreted** version of "jump"
  - On arrival of a customer, the entire load **dispersed** over her lifetime **can** be expressed as a jump at her arrival.
  - **Mathematically feasible**: only stationarity! (cf., $Z_n = R_n$ in Little's Formula)

- **Contributions** of Palm Calculus
  - **Extension** of time average (or event average expression)
  - An **edifice of notations** – mathematical condensation/compactification
  - Concise rephrasing indeed helps **clarify** how you think and view a system

4. PASTA

- There is an important case where Event average = Time average
- "**P**oisson **A**rrivals **S**ee **T**ime **A**verages"
  - More exactly, it should be:

Poisson Arrivals *independent of simulation state* See Time Averages

Consider a system that can be modeled by a stationary Markov chain $S(t)$ in discrete or continuous time (in practice any simulation that has a stationary regime and is run long enough). We are interested in a matrix of $C \geq 0$ of selected transitions such that

Independence For any state i of $S(t)$, $\sum_j C_{i,j} \stackrel{\text{def}}{=} \lambda$ is independent of i.

THEOREM 7.5.2 (PASTA). *Consider a point process of selected transitions as defined above. The Palm probability just before a transition is the stationary probability.*

A simplified version of which is "(Homogeneous) Poisson process"

EXAMPLE 7.19: A POISSON PROCESS THAT DOES NOT SATISFY PASTA. The PASTA theorem requires the event process to be Poisson or Bernoulli *and* independence on the current state. Here is an example of Poisson process that does not satisfy this assumption, and does not enjoy the PASTA property.

Construct a simulation as follows. Requests arrive as a Poisson process of rate λ, into a single server queue. Let T_n be the arrival time of the nth request. The service time of the nth request is assumed to be $\frac{1}{2}(T_{n+1} - T_n)$. The service times are thus exponential with mean $\frac{1}{2\lambda}$, but not independent of the arrival process. Assuming the system is initially empty, there is exactly 1 customer during half of the time, and 0 customer otherwise. Thus the time average distribution of queue length $X(t)$ is given by $\mathbb{P}(X(t) = 0) = \mathbb{P}(X(t) = 1) = 0.5$ and $\mathbb{P}(X(t) = k) = 0$ for $k \geq 2$. In contrast, the queue is always empty when a customer arrives. Thus the Palm distribution of queue length just before an arrival is different from the time average distribution of queue length.

The arrival process does not satisfy the independence assumption: at a time t where the queue is not empty, we know that there cannot be an arrival; thus the probability that an arrival occurs during a short time slot depends on the global state of the system.

Service time depends on T_{n+1}, i.e., the future arrival!
Which means all arrival events depend on the system.

EXAMPLE 7.17: ARP REQUESTS WITHOUT REFRESHES. IP packets delivered by a host are produced according to a Poisson process with λ packets per second in average. When a packet is delivered, if an ARP request was emitted not more than t_a seconds ago, no ARP request is generated. Else, an ARP request is generated. What is the rate of generation of ARP requests ?

Call T_n the point process of ARP request generations, μ its intensity and p the probability that an arriving packet causes an ARP request to be sent. First, we have $\mu = p\lambda$ (to see why, assume time is discrete and apply the definition of intensity).

Second, let $Z(t) = 1$ if the ARP timer is running, 0 if it has expired. Thus p is the probability that an arriving packet sees $Z(t) = 0$. The PASTA property applies, as the IP packet generation process is independent of the state of the ARP timer.

By the inversion formula:

Time average — Conditional on the event of ARP request generation

$$p = \mathbb{P}(Z(t) = 0) = \mu \mathbb{E}^0(T_1 - t_a) = \mu\left(\frac{1}{\mu} - t_a\right) = 1 - \mu t_a \tag{7.56}$$

∵PASTA

Combining with $\mu = p\lambda$ gives $p = \frac{1}{\lambda t_a + 1}$, and the rate of generation of ARP requests is $\mu = \frac{\lambda}{1+\lambda t_a}$.

$$p = \mathrm{P}(Z(t) = 0) = \mathrm{E}\left(1_{\{Z(t)=0\}}\right) = \mu \mathrm{E}^0\left(\int_0^{T_1} 1_{\{Z(t)=0\}}\, dt\right) = \mu \mathrm{E}^0\,(T_1 - t_a)$$

5. RWP and Freezing Simulations

- Modulator Model (Supplement of mathematical details to Stochastic Occurrence.)

Recall that a stochastic recurrence is defined by a sequence Z_n, $n \in \mathbb{Z}$, (also called the modulator state at the nth epoch) and a sequence $S_n > 0$, interpreted as the duration of the nth epoch. The state space for Z_n is arbitrary, not necessarily finite or even enumerable. We assume that (Z_n, S_n) is random, but stationary[2] with respect to the index n. As usual, we do not assume any form of independence.

Roughly, it's merely a continuous version.

We are interested in the *modulated process* $(Z(t), S(t))$ defined by $Z(t) = Z_n, S(t) = S_n$ whenever t belongs to the nth epoch (i.e. when $T_n \le t < T_{n+1}$). We would like to apply Palm calculus to $(Z(t), S(t))$.

EXAMPLE 7.13: LOSS CHANNEL MODEL. A path on the internet is modelled as a loss system, where the packet loss ratio at time t, $p(t)$ depends on a hidden state $Z(t) \in \{1, ..., I\}$ (called the modulator state). During one epoch, the modulator remains in some fixed state, say i, and the packet loss ratio is constant, say p_i. At the end of an epoch, the modulator changes state and a new epoch starts.

Once in a while we send a probe packet on this path, thus we measure the time average loss ratio $\bar{p}$. How does it relate to p_i ? Apply the inversion formula:

$$\bar{p} = \frac{\sum_i \pi_i^0 p_i \bar{S}_i}{\sum_i \pi_i^0 \bar{S}_i}$$

Integral of "loss ratio" between two epochs (numerator)

$1/\lambda$ (denominator)

where π_i^0 is the probability that the modulator is in state i at an arbitrary epoch (proportion of i epochs) and $\bar{S}_i$ is the average duration of an i-epoch.

Is the previous simulation stationary ?

- Seems like a superfluous question, however there is a difference in viewpoint between the epoch n and time.

- Let $\boldsymbol{S_n}$ **be the length of the** $\boldsymbol{n^{\text{th}}}$ **epoch**.
- If there is a stationary regime, then by the inversion formula

$$\lambda = \frac{1}{\int_0^\infty t f_S^0(t)dt} > 0$$

so the mean of S_n must be finite.

- This is in fact **sufficient** (and necessary)

THEOREM 7.4.1. *Assume that the sequence* S_n ~~*satisfies* ***H1*** *and*~~ *has finite expectation. There exists a stationary process* $Z(t)$ *and a stationary point process* T_n *such that*

1. $T_{n+1} - T_n = S_n$
2. $Z_n = Z(T_n)$

Finite expectation of epoch ➔ Stationarity

Application to RWP

EXAMPLE 7.14: RANDOM WAYPOINT, CONTINUATION OF EXAMPLE 7.6. For the random waypoint model, the sequence of modulator states is

$$Z_n = (M_n, M_{n+1}, V_n)$$

and the duration of the nth epoch is

$$S_n = \frac{d(M_n, M_{n+1})}{V_n} \tag{7.48}$$

where $d(M_n, M_{n+1})$ is the distance from M_n to M_{n+1}.

Can this be assumed to come from a stationary process ? We apply Theorem 7.4.1. The average epoch time is

$$\mathbb{E}(S_0) = \mathbb{E}\left(\frac{d(M_n, M_{n+1})}{V_n}\right) = \mathbb{E}\left(d(M_n, M_{n+1})\right)\mathbb{E}\left(\frac{1}{V_n}\right)$$

since the waypoints and the speed are chosen independently. Thus we need that $\mathbb{E}\left(\frac{1}{V_n}\right) < \infty$, i.e. $v_{\min} > 0$.

A Random waypoint model that has no stationary regime !

- Assume that at trip transitions, node speed is sampled uniformly on $[v_{\text{min}}, v_{\text{max}}]$
- Take $v_{\text{min}} = 0$ and $v_{\text{max}} > 0$
- Mean trip duration = (mean trip distance) $\times \frac{1}{v_{\text{max}}} \int_0^{v_{\text{max}}} \frac{dv}{v} = +\infty$

  $\mathbb{E}(S_0)$ $\mathbb{E}\left(d\left(M_n, M_{n+1}\right)\right)$ $\mathbb{E}\left(\frac{1}{V_n}\right)$
- **Mean trip duration is infinite!**
- **Very often used in a number of research papers**
- Speed decay: “considered harmful” [YLN03]
  - It took **a couple of decades** for us to be enlightened.
  - **Exclusion of zero speed**, e.g., $(0, v_{\text{max}}]$, still results in infinite mean drip duration

[YLN03] J. Yoon, M. Liu, and B. Noble, “Random Waypoint Considered Harmful”, *IEEE Infocom, 2003*.

What happens when the model does not have a stationary regime ?

- When $v_{\text{min}} = 0$, the simulation becomes old.
- Also, the sample average speed decays to 0.

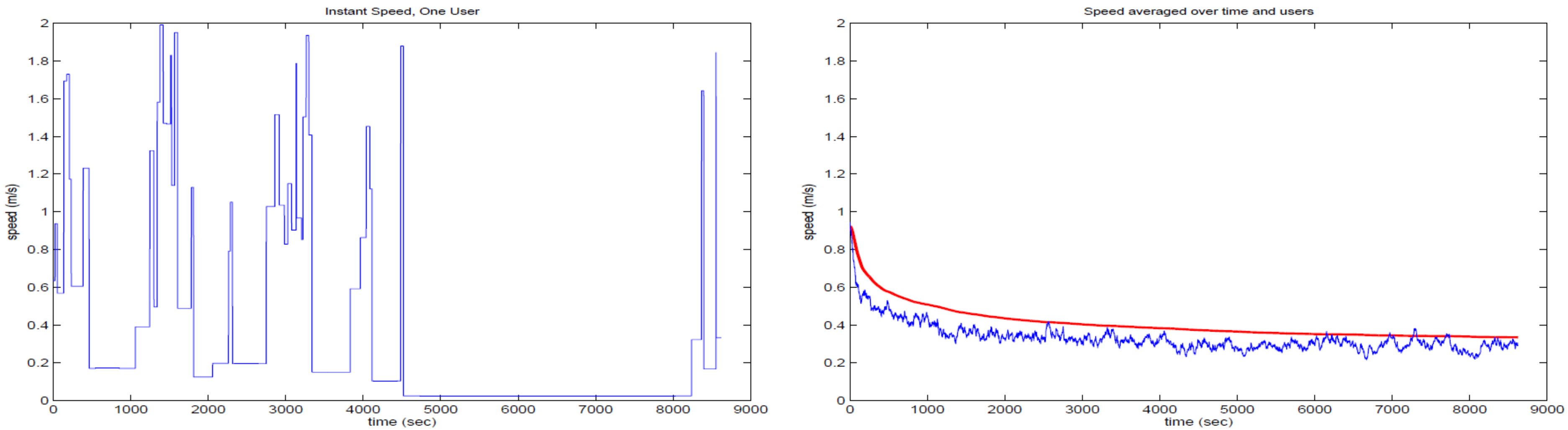


Figure 7.9: Freezing simulation: random waypoint with $v_{\min} = 0$. The model does not have a stationary regime and the simulation becomes slower and slower. First panel: sample of instant speed versus time for one mobile. Second panel: speed averaged over $[0; t]$ for one mobile (zig zag curve) or for 30 mobiles (smoother curve). The average speed slowly tends to 0.

You may want to marginalize this finding as a mathematical trivia. However, if you think hard, it's not in line with your intuition because exclusion of $v_{\text{min}} = 0$ does not mitigate the situation.

Stationary Distribution of Speed
(For model with stationary regime)

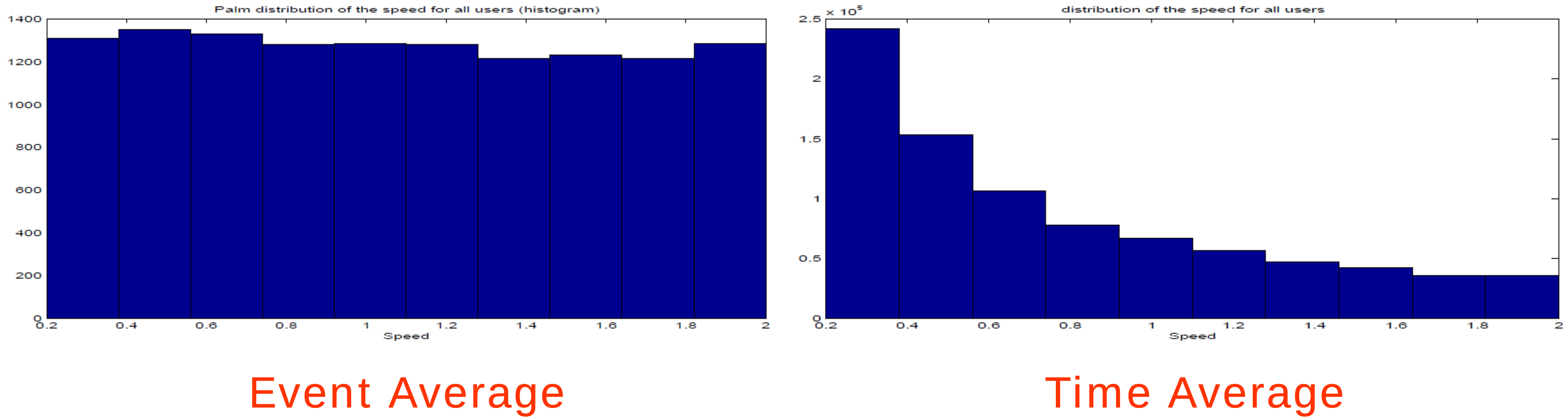


Figure 7.4: Distribution of speed sampled at waypoint (first panel) and at an arbitrary time instant (second panel). $v_{\min} = 0.2, v_{\max} = 2$m/s.

Easy to analyze with Palm theory.
Wish if they had understood it through the lens of Palm theory!

Closed Form

- Assume a stationary regime exists and simulation is run long enough
- Apply inversion formula and obtain distribution of instantaneous speed V(t)

$$\mathbb{E}\left(\phi(V(t))\right) = \lambda \mathbb{E}^0 \left(\int_0^{T_1} \phi(V(t))\, dt \right)$$

Time average of a function of $V(t)$

$$\begin{aligned}
&= \lambda \mathbb{E}^0 \left(\phi(V_0) T_1\right) \\
&= \lambda \mathbb{E}^0 \left(\phi(V_0) \frac{\|M_1 - M_0\|}{V_0} \right) \\
&= \lambda \mathbb{E}^0 \left(\|M_1 - M_0\| \right) \mathbb{E}^0 \left(\frac{\phi(V_0)}{V_0} \right) \\
&= C \int_0^{v_{\max}} \frac{\phi(v)}{v} f^0_{V_0}(v) dv
\end{aligned}$$

C

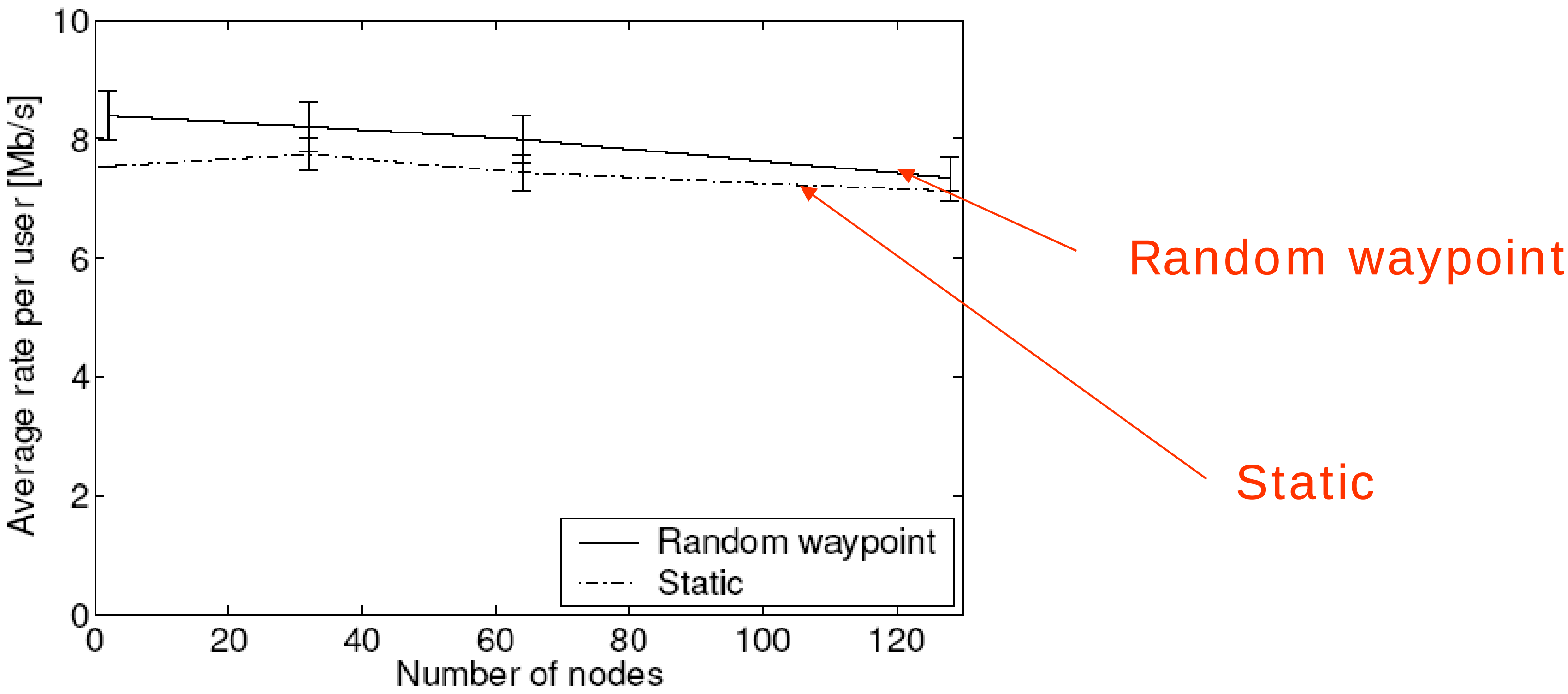
$$f_{V(t)}(v) dv = \frac{C}{v} f^0_{V_0}(v) dv$$

"Real" distribution of $V(t)$ is lighter-tailed than its Palm version.

Removing Transient Matters

- A (true) example: Compare impact of mobility on a protocol:
  - Experimenter places nodes uniformly for static case, according to random waypoint for mobile case
  - Finds that static is better
- **Q.** Find the bug !

- **A.** In the mobile case, the nodes are more often **towards the center**, distance between nodes is **shorter**, performance is **better**
- The comparison is flawed. Should use for static case the same distribution of node location as random waypoint.



Curiosity

- Is it possible to have the **time distribution** (as opposed to Palm distribution) of speed uniformly distributed in $[0, v_{\max}]$?

Yes. Use the inversion formula $f_{V(t)}(v)dv = \frac{C}{v} f^0_{V_0}(v)dv$

At a waypoint, pick a new speed according to the distribution

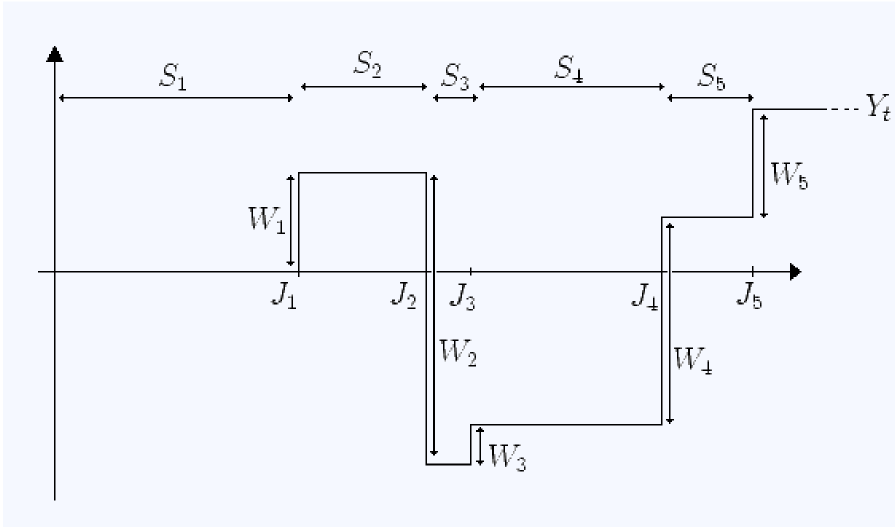
$$f^0_V(v) = Kv1_{0 \le v \le v_{\max}}$$

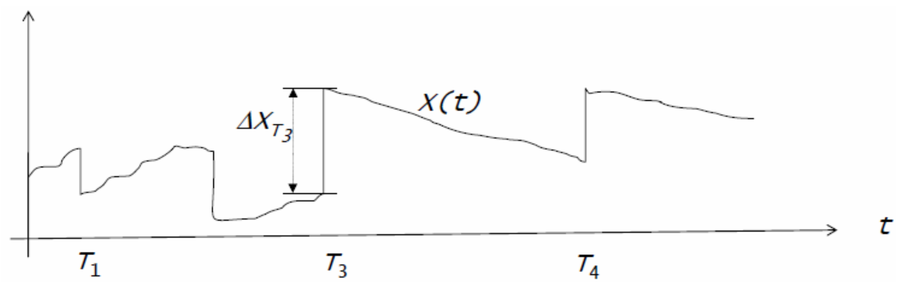
$$\therefore \mathbb{E}\left(\frac{1}{V_n}\right) < \infty$$

6. Application to Throughput Analysis

- Techniques for Throughput Analysis
  - Renewal Reward Theorem

  (also applicable to Markov chains)
  - Palm Inversion Formula
- Renewal Reward Theorem
  - Observations or metrics (a.k.a. 'Rewards') W_n are **independent and identically distributed** (iid)
  - Inter-transition times (a.k.a. 'Renewals') S_n are **independent and identically distributed** (iid)
- Palm Inversion Formula
  - Observations or metrics X_t, and inter-transition times $T_n - T_{n-1}$ are only required to be jointly stationary
  - **All kinds of dependencies are allowed.**





Throughput Formula in IEEE 802.11 MAC

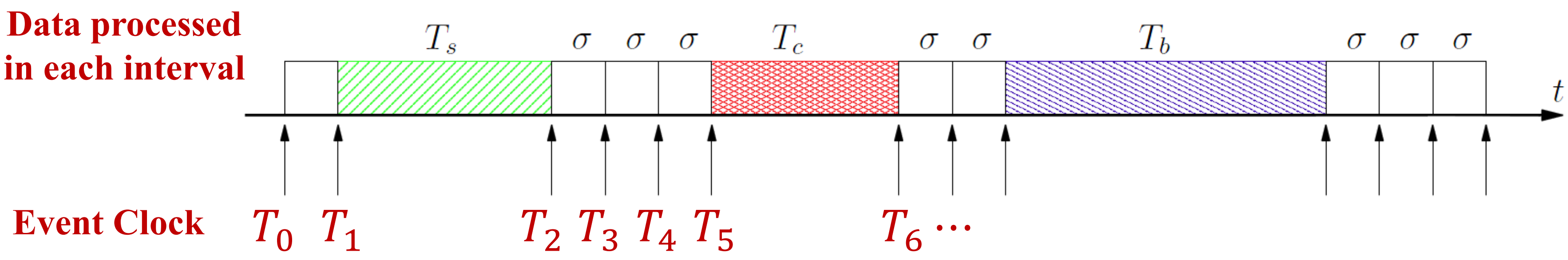


Fig. 1. The channel view of a node

- The time intervals during which the node (of interest) remains in each of the four states, (i) idle channel; (ii) channel occupied by a successful transmission of the node; (iii) channel occupied by a collision of the node; (iv) busy channel due to activity of other nodes are respectively denoted by:

$$\sigma, T_s, T_c, T_b$$

- The probabilities that (i) the node sends out a packet after an idle slot; (ii) a transmission of the node is not successful; (iii) the channel becomes busy after an idle slot due to activity of other nodes are respectively denoted by:

$$\tau, p, b$$

With complicated probability distributions

Throughput Formula in IEEE 802.11 MAC

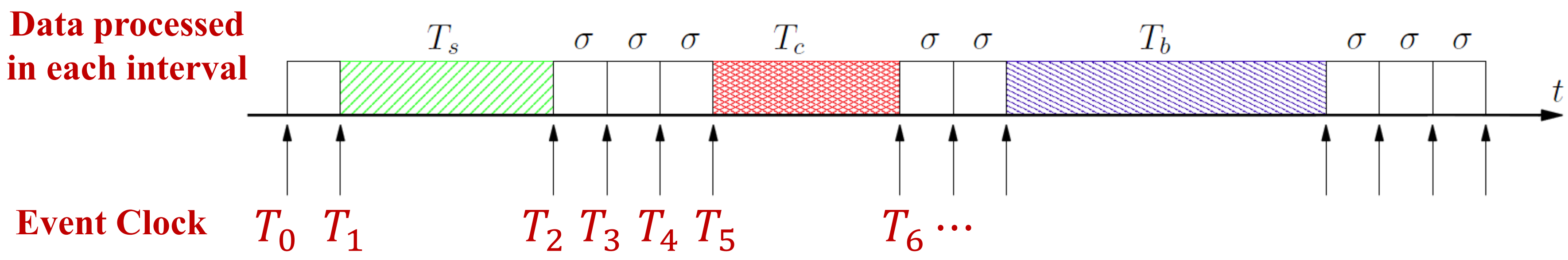


Fig. 1. The channel view of a node

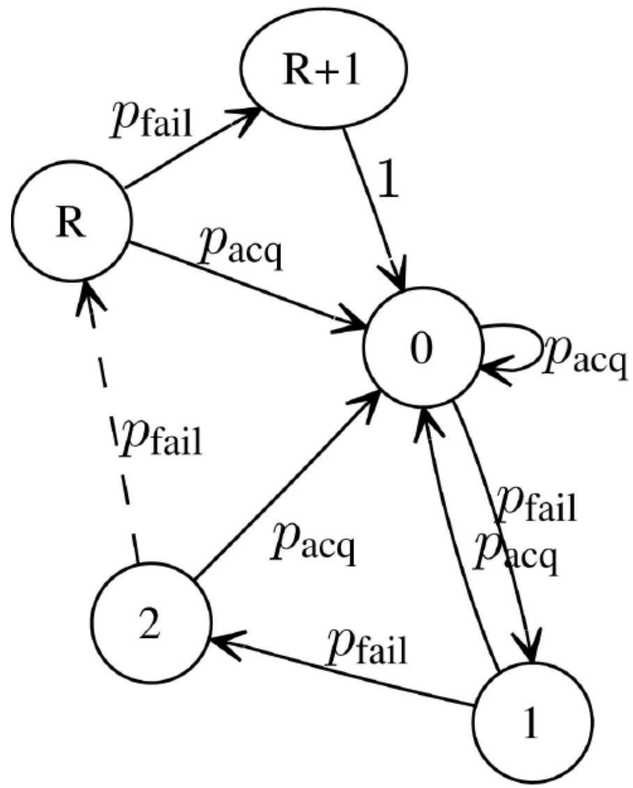
$$T_P = \frac{\tau(1-p)}{\tau(1-p)\bar{T}_s + \tau p \bar{T}_c + (1-\tau)(1-b)\sigma + (1-\tau)b\bar{T}_b}$$

By Palm Inversion formula

[GSK08] M. Garetto, T. Salonidis, E. Knightly, "Modeling Per-flow Throughput and Capturing Starvation in CSMA Multi-hop Wirelss Networks", *IEEE/ACM Trans. Networking*, 2008.

Throughput Formula in IR-UWB Networks

- Correct packet reception in IR-UWB (Impulse Radio Ultra-wideband) networks
  - Packet detection
  - Timing acquisition
  - Retransmissions



Saturation throughput in packets per second

$$\lambda_0 = \frac{p_{\text{acq}}\left(1-\pi_X(R+1)\right)}{p_{\text{acq}}(t_{\text{acq}}+t_{\text{tx}})\left(1-\pi_X(R+1)\right)+p_{\text{fail}}\sum_{i=0}^{R}\left(t_{\text{acq}}+t_{\text{fail}}(i)\right)\pi_X(i)+t_{\text{drop}}\pi_X(R+1)}$$

Palm calculus gets rid of *exponential inter-transition time* assumption in Markov chains.

[MER09] R. Merz and J.-Y. Le Boudec, "Performance Evaluation of Impulse Radio UWB Networks Using Common or Private Acquisition Preambles", *IEEE Trans. Mobile Computing*, 2009.

Peppering Your Formal Analysis

- Energy-efficient Wi-Fi sensing algorithms for an arbitrary inter-AP time distribution

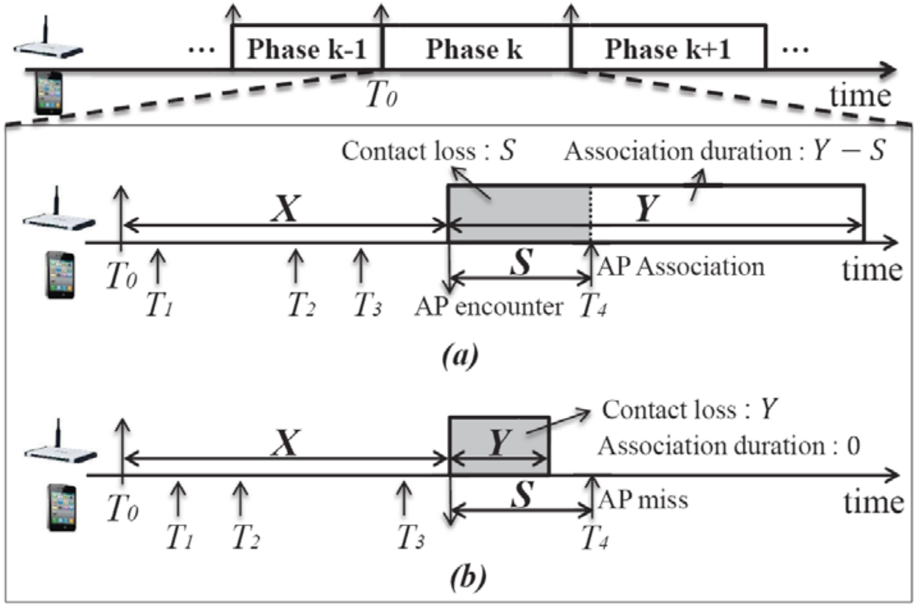
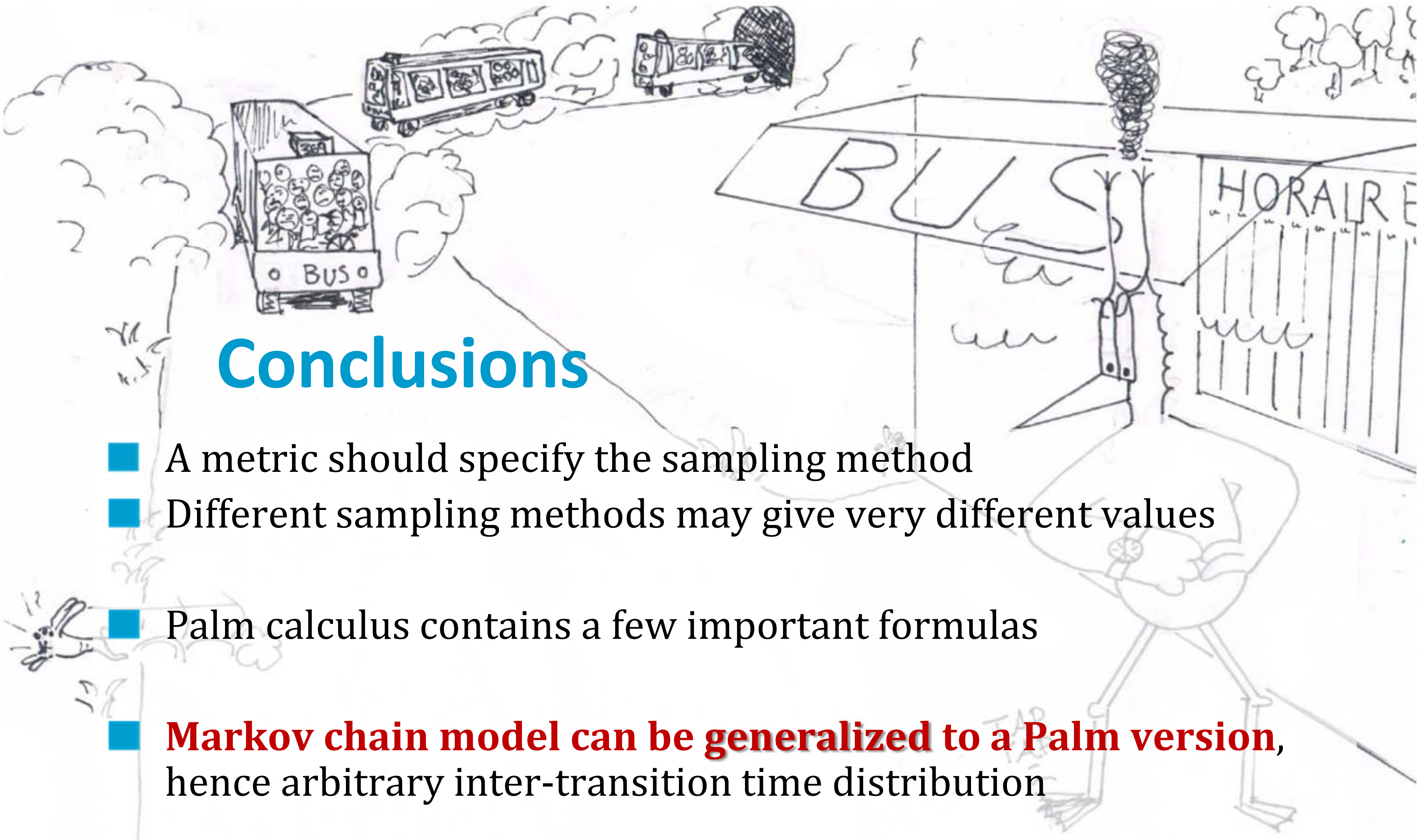


Fig. 1. An illustration of the model. (a) When AP contact duration Y is larger than S, a mobile successfully detects an AP. (b) When AP contact duration Y is smaller than S, a mobile node fails to detect it. Shaded areas denote the contact loss time $l(S)$.

- **Objective**: Minimize energy for sensing (T_1,T_2,...) plus contact loos time (S)
- **Residual time until the next AP** ($I(t)$ is duration of current interval):

$$\mathrm{E}[S] = \frac{1}{2}\left(\mathrm{E}[I(t)] + \frac{\left(\sigma_{I(t)}\right)^2}{\mathrm{E}[I(t)]}\right)$$

[JEO13] J. Jeong, Y. Yi, J. Cho, D. Eun and S. Chong, "Wi-Fi Sensing: Should Mobiles Sleep Longer As They Age?", *IEEE Infocom*, 2013.



Conclusions

- A metric should specify the sampling method
- Different sampling methods may give very different values

- Palm calculus contains a few important formulas

- **Markov chain model can be generalized to a Palm version**, hence arbitrary inter-transition time distribution

- Yet the most rational stance is to view Palm as a solid **intermediary** and **mathematical language** for advanced skills.