PRM: "Modelado de Tráfico"

3º Ingeniería de Telecomunicaciones — UPV/EHU

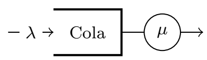
"Under-promise and over-deliver."

Javier de Martín – 2016/17

Variables

λ: Tasa de llegada de usuarios (nº usuarios llegados / tiempo)
$1/\lambda$: Tiempo medio entre llegadas consecutivas
μ: Tasa de servicio
$1/\mu$: Tiempo medio de servicio
C: Capacidad del canal
$\rho = \frac{\lambda}{\mu}$: Factor de utilización (intensidad de tráfico)
N: Estado del sistema, número de clientes en el sistema
P_B: Probabilidad de bloqueo.
γ: Throughput

Fundamentos de la Teoría de Colas



Cuando $\lambda \to \mu$ la cola se hace inestable, para que sea **estable**: $\lambda < \mu$.
Si $\rho \to 1$ los tiempos de espera aumentan y la cola se llena rápidamente.
El **estado de la cola** es el número de usuarios/paquetes en la cola incluido el que pueda estar en el servidor.
Un **proceso de Poisson** es un caso particular del proceso de Markov en el cual la probabilidad de ocurrencia de un evento en $t + \Delta t$ es independiente de sucesos anteriores.
En una **distribución de Poisson** la probabilidad de k llegadas es:

$$p(k) = \frac{(\lambda \cdot T)^k}{k!} \cdot e^{-\lambda T} \quad \rightarrow \quad \sum_{k=0}^{\infty} p(k) = 1$$

La **esperanza** es $E(k) = \lambda \cdot T$ y **varianza** $\sigma^2(k) = \lambda \cdot T$.
Tiempo entre eventos sucesivos, función de densidad de probabilidad:

$$f_\tau(\tau) = \lambda e^{-\lambda\tau} = \frac{\partial F_\tau(x)}{\partial x}$$

Relación de los procesos exponenciales con los poissonianos, probabilidad de una llegada en un intervalo:

$$Pr\{\tau > x\} = p_x(0) = e^{-\lambda x}$$

Función de distribución de probabilidad acumulativa, mide la probabilidad de que τ sea menor o igual a x:

$$F_\tau(x) = 1 - e^{\lambda x}$$

Distribución del tiempo de servicio:

$$f_r(r) = \mu \cdot e^{-\mu r}$$

Propiedad: Suponer m procesos de Poisson independientes de tasas λ_1, λ_2, ..., λ_m que se combinan, el proceso resultante será también de Poisson con una tasa $\lambda = \sum_{i=1}^{m} \lambda_i$

Notación Kendall

$$A/B/m/K/c/z$$

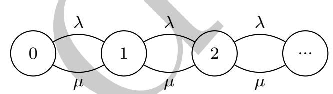
A: Distribución del tiempo entre llegadas **B**: Distribución del tiempo de servicio **m**: Número de servidores en paralelo en el sistema **K**: Capacidad de la cola (incluyendo servidores) **c**: Tamaño de la fuente **z**: Disciplina de la cola
Las distribuciones de llegadas y servicio pueden ser:

- **M**: Exponencial
- E_k: Erlang de k etapas
- H_k: Hiperexponencial de k etapas
- **G**: General
- **D**: Determinística (constante)

Por defecto para colas A/B/N la cola es infinita y la disciplina FIFO.
Cuando el sistema alcanza su estado más estable se supone que la probabilidad de usuarios en el mismo no varía con el tiempo. Con el tiempo los valores de p_n se irán aproximando a los valores más estables.

M/M/1

Las **ecuaciones de balance** permiten calcular la probabilidad de estado p_n. En el diagrama de estados, el # es el número de clientes en el sistema. El factor de utilización es $\rho = \lambda/\mu$.



$$\begin{cases} (n-1) & \rightarrow \text{Ha llegado un usuario} \\ n & \rightarrow \text{No ha llegado ni salido nadie} \\ (n+1) & \rightarrow \text{Ha salido un usuario} \end{cases}$$

$$\lambda \cdot p_n = \mu \cdot p_{n+1} \quad n = 0, 1, 2... \qquad \sum_{n=0}^{\infty} p_n = 1$$

$$p_{n+1} = \rho \cdot p_n$$

Es condición necesaria que $\rho < 1$:

$$p_0 = \frac{1}{\sum_{n=0}^{\infty} p_n = 1} = (1-\rho) \qquad p_n = (1-\rho)\cdot\rho^n \qquad \forall\rho = \frac{\lambda}{\mu}$$

Es condición necesaria que $\rho < 1$:

$$p_0 = \frac{1-\rho}{1-\rho^{N+1}} \qquad p_n = \frac{1-\rho}{1-\rho^{N+1}} \cdot \rho^n \qquad \forall\rho = \frac{\lambda}{\mu}$$

La **probabilidad de bloqueo** es $P_N = p_n$.

Tiempo medio por cliente en el sistema:

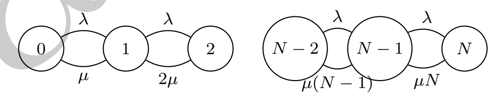
$$\frac{1}{\mu(1-\rho)}$$

Tiempo medio de espera en cola:

$$\frac{\rho}{\mu(1-\rho)}$$

Cola M/M/N

Permite procesamiento en paralelo al tener múltiples servidores, el factor de utilización es $\rho = \lambda/m\mu$.
La tasa media de llegada λ es constante e independiente del estado del sistema ($\lambda_n = \lambda$).



$$p_n = \frac{n+1}{N\rho} \cdot p_{n+1} \qquad n = 0, ..., N-1$$

$$p_{n+1} = \rho \cdot p_n \qquad n \geq N$$

Throughput (γ)

$$\gamma = \lambda \cdot (1 - P_B)$$

$$= \mu \cdot (1 - p_0$$

Para una **cola infinita**:

$$\gamma = \mu \cdot \rho = \lambda$$

Para una **cola finita**:

$$P_B = \frac{1-\rho}{1-\rho^{N+1}} \cdot \rho^N = P_N$$

Cola M/M/1/N en función de ρ

Si el sistema está lleno, N, no se permite la entrada a nuevos clientes al sistema. La tasa de llegadas no es constante y varía con el tiempo en función de si el sistema está lleno o no:

$$\lambda_{ef} = \lambda(1 - p_k)$$

$$p_n = \rho^n \cdot p_0$$

Para estos sistemas no existe el estado $k+1$.

$$p_n = \frac{1-\rho}{1-\rho^{N+1}} \cdot \rho^n$$

Región de congestión

Throughput Normalizado (γ/μ)

Mide la relación entre los usuarios que son atendidos en el sistema por unidad de tiempo y los que potencialmente podrían haber sido atendidos.

$$\frac{\gamma}{\mu} = \frac{1-\rho^N}{1-\rho^{N+1}} \cdot \rho = \frac{Usuariosatendidos/seg}{Usuariospotenciales/seg}$$

Insertar Gráfica

Número medio de usuarios en la cola M/M/1 E(n)

$$E(n) = \frac{\rho}{1-\rho}$$

Tiempo Medio de Estancia en la Cola E(T)

$$E(T) = \frac{E(n)}{\gamma} \qquad \rightarrow \qquad E(T)_{M/M/1} = \frac{1}{(1-\rho)\cdot\mu} = \frac{1}{\mu-\lambda}$$

Tiempo Medio de Espera en la Cola E(w)

$$E(w) = E(T) - 1/\mu$$

Número medio de usuarios esperando en la cola E(q): $E(q) = \gamma \cdot E(w)$

Colas dependientes del estado: Proceso de nacimiento y muerte

En las **colas dependientes del estado** las tasas de llegada λ y las de servicio μ dependen del estado del sistema, presentes en colas de servidores múltiples: $M/M/m$ y .

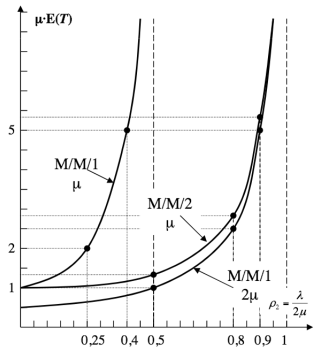
$$\lambda = \lambda(n) = \lambda_n \rightarrow \text{ Procesos de nacimiento}$$
$$\mu = \mu(n) = \mu_n \rightarrow \text{ Procesos de muerte}$$

Un **proceso de nacimiento** como un caso de procesos de Markov donde las tasas de llegada son función del estado. Un **proceso de muerte** es otro caso de procesos de Markov donde también las salidas son función del estado de la cola.

Ecuaciones de Balance

$$p_n = \frac{\prod_{i=0}^{n-1}\lambda_i}{\prod_{i=1}^{n}\mu_i}\cdot p_0 \qquad p_0 = \frac{1}{\sum_{n=1}^{N}\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}+1}$$

[1] Envío bidireccional pero no simultáneo



Fórmulas de Pollaczek Khinchine para distribuciones M/D/1:
Número medio de usuarios E(n):

$$E(n) = \frac{\rho}{1-\rho}\cdot\left(1-\frac{\rho}{2}\right)$$

Tiempo medio de estancia en la cola E(T):

$$E(T) = \frac{E(n)}{\gamma} = \frac{1}{\mu(1-\rho)}\cdot\left(1-\frac{\rho}{2}\right)$$

Modelado de Protocolos de N.Enlace

"Su función principal es el control del enlace de datos entre dos nodos adyacentes."

Stop & Wait

Suposiciones para el estudio analítico:

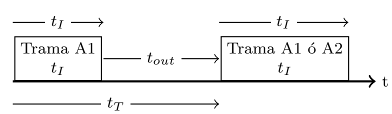
Sólo se permite transmisión de datos `A` → `B`, `B` sólo envía `ACK` o `NACK`.

`A` transmite continuamente → saturación.

El cálculo del throughput se realiza utilizando el límite máximo.

Envía cada trama y espera recibir un ACK o NACK:

- Si llega un ACK se libera el buffer que almacenaba el paquete transmitido y se envía otra trama.
- Si llega un NACK o expira el temporizador, se retransmite la trama.



nº	Tiempo	Probabilidad
1	$t_T = t_I + t_{out}$	$1-p$
2	$2\cdot t_T$	$p(1-p)$
...	...	...
n	$n\cdot t_T$	$p^{n-1}(1-p)$

- t_I: Tiempo requerido para transmitir una trama
- t_{out}: Tiempo del temporizador de espera de ACK
  $t_{out} \geq 2\cdot t_p + t_{proc} + t_s$
  - t_p: Tiempo de retardo de propagación
  - t_{proc}: Tiempo de procesado en recepción
  - t_s: Tiempo de transmisión de respuesta
- t_T: Tiempo mínimo entre tramas de datos sucesivas
  $t_T = t_I + t_{out}$

Apropiado para la transmisión semiduplex [1]. No es apropiado para transmisión duplex si el tiempo de propagación es mucho mayor que el tiempo empleado para transmitir la trama.
Debido a retransmisiones provocadas por tramas erróneas:

$$\gamma_{real} < \gamma_{max}$$

Tiempo medio para una transmisión correcta:

$$t_v = t_T(1-p) + 2t_T(1-p)p + 3t_T(1-p)p^2 + ... =$$
$$= t_T(1-p)\sum_{i=1}^{\infty} i\cdot p^{i-1} = t_T(1-p)\frac{1}{(1-p)^2} =$$
$$= \frac{t_T}{1-p}$$

Throughput máximo ($TH_{real} < TH_{max} \leftarrow$ debido a retransmisiones provocadas por tramas erróneas)

$$TH_{max_{S\&W}} = \lambda_{max_{S\&W}} = \frac{1}{t_v} = \frac{1-p}{t_T}$$
$$= \frac{1-p}{a\cdot t_I}(paq/seg) \qquad a = \frac{t_T}{t_I} \geq 1$$

Throughput normalizado real ρ:

$$\rho \equiv \lambda\cdot t_I \leq \lambda_{max}\cdot t_I = \frac{1-p}{a} < 1$$

Go-Back-N

Envía continuamente las tramas, mejorando el throughput. Es útil para comunicaciones full-duplex.

Se realizan las siguientes **suposiciones**:

Los números de secuencia se consideran indefinidos.

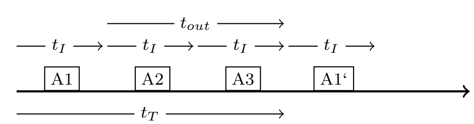
La retransmisión por `NACK` o t_{out} se estudian igual.

Transmisión en régimen de saturación de `A` → `B`.

Longitud de trama t_I fija.

Temporizador t_{out} fijo y es múltiplo de t_I.

Mínimo tiempo entre transmisiones es el tiempo de transmisión de una trama t_I.



Transmisión con errores

nº	Tiempo	Probabilidad
1	t_I	$1-p$
2	$t_I + t_T$	$p(1-p)$
...	...	...
n	$t_I + (n-1)\cdot t_T$	$p^{n-1}(1-p)$

Tiempo medio de transmisión de una trama:

$$t_V = (1-p)\cdot t_I + (1-p)\cdot p(t_T + t_I) + (1-p)\cdot p^2(2t_T + \cdot t_I) + ... =$$
$$= (1-p)\cdot t_I + (1-p)t_T \sum_{i=1}^{\infty} i \cdot p^i + (1-p)\cdot t_I \sum_{i=1}^{\infty} p^i =$$
$$= (1-p)\cdot t_I + a \cdot t_I \cdot (1-p)\frac{p}{(1-p)^2} + t_I \cdot (1-p)\cdot \frac{p}{1-p} =$$
$$= t_I(1-p+\frac{ap}{1-p}+p) = t_I \frac{1-p+ap}{1-p} =$$
$$= t_I \frac{1+(a-1)p}{1-p} \qquad a \equiv \frac{t_T}{t_I} = \frac{t_I + t_{out}}{t_I} = 1 + \frac{t_{out}}{t_I}$$

Throughput máximo:

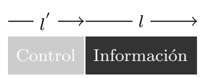
$$\lambda_{max} = \frac{1}{t_v} = \frac{1-p}{t_I \cdot (1+(a-1)p)}$$

Throughput normalizado:

$$\rho = \lambda \cdot t_I \leq \frac{1-p}{1+(a-1)p} < 1$$

$$\lambda_{max_{GBN}} = a \cdot \lambda_{max_{S\&W}}$$

ACKs embebidos en tramas B-A
Influencia de la tasa de error



Erorres en enlaces de satélite, como **hipótesis** se toma que p_b es la probabilidad de error de bit independiente de la posición del bit debido al ruido aleatorio. Probabilidad de error de trama p: $p = 1-(1-p_b)^{l+l'} = 1 - q_b^{l+l'}$. Si $p_b << 1 \rightarrow p \approx (l+l')p_b \rightarrow p << 1$.
Errores en enlaces terrestres, los errores se producen en ráfagas. p es proporcional a $l+l'$, si $p_b << 1 \rightarrow p \approx (l+l')k$.
Tasa de datos normalizada D/C, la estación emisora está en saturación (λ_{max}) y se utiliza el protocolo tipo Go-Back-N.

$$D = \lambda_{max} \cdot l = l \cdot \frac{1-p}{t_I \cdot (1+(a-1))p} \qquad t_I = \frac{l+l'}{C}$$

$C \equiv$ Capacidad del canal

$$\frac{D}{C} = \frac{l}{l+l'} \cdot \frac{1-p}{t_I \cdot (1+(a-1))p} \leq 1$$

$$\frac{D}{C} \approx \frac{l}{l+l'} \cdot (1-p)$$

Cálculo de la longitud óptima

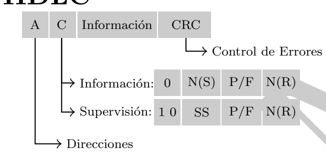
Hipótesis: $a = 1$ (Stop & Wait) o $(a-1)\cdot p \ll 1$.
Caso general de **satélite**:

$$l_{opt} = \frac{l'}{2} \cdot \left(\sqrt{1 - \frac{4}{l' \cdot L_n q_b}} - 1\right)$$
$$\approx \sqrt{\frac{l'}{p_b}} \quad \text{suponiendo } p_b \ll 1 \rightarrow p_b \cdot l' \ll 1$$

Caso **terrestre/satélite** simplificado $p = p_b(l+l')$:

$$l_{opt} =\approx \sqrt{\frac{l'}{p_b}}$$

Protocolo HDLC



SS → tipos de tramas de supervisión:

- RR (*Ready to Receive*): Confirma las tramas hasta $N(R)-1$.
- RNR (*Not Ready to Receive*): Confirma las tramas hasta $N(R)-1$ y establece control de flujo con condición de bloqueo.
- REJ (*Reject*): Confirma las tramas hasta $N(R)-1$ y retransmisión desde la trama $N(R)$.
- SREJ (*Selective Reject*): Pide retransmisión de la trama $N(R)$.

Análisis en HDLC/ABM:
Se **diferencia respecto a Go-Back-N** de que tiene numeración de secuencia finita y procedimiento de control de errores específico.
Procedimiento de control de errores, se define una implementación específica sobre la cual se realiza el análisis del protocolo:

Recuperación vía REJ siempre que se pueda acelerar el procedimiento de recuperación. Sólo una vez por la misma trama.

Recuperación por temporizador (además de REJ) para las retransmisiones sucesivas y errores no detectados por REJ.

Envío de tramas P/F para obligar a respuestas forzadas en situaciones especiales de recuperación.

Hipótesis para el análisis:
Sólo se transmite en el sentido $A \rightarrow B$.
A transmite en el régimen de saturación.
B responde con RR (ACK) y REJ $(NACK)$.
Datos para el análisis:
Módulo: Número de tramas con diferente numeración que pueden ser enviadas.
Ventana: Número máximo de tramas que pueden estar en circulación sin haber sido confirmadas.

Longitud de tramas: Información (t_I) y de supervisión (t_S):

$$t_I = \frac{l+l'}{C} \qquad t_S = \frac{l'}{C}$$

Retardo de propagación (incluyendo t_{proc}): t_p
Tiempo de confirmación: $t_{ack} = 2\cdot t_p + t_s$
Temporizador de reenvío: $t_{out} = 2 \cdot t_p + 2 \cdot t_I > t_{ack}$
TERMINAR
Cálculo del Throughput la **hipótesis** es que cuando ocurre un error se puede recuperar mediante: `REJ` (primer caso) o `TEMPORIZADOR` (siguientes casos).
Throughput máximo:

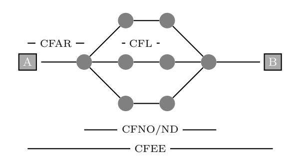
$$\lambda_{max} = \frac{1}{t_v} = \frac{1-p}{t_I(1+(a-1)p)}$$

p: Probabilidad de recibir una trama con error
T_1: Tiempo aleatorio requerido para la primera retransmisión (puede tomar diferentes valores)
T_2: Tiempo medio requerido para retransmisiones posteriores (es siempre igual en cada transmisión)
t_v: Tiempo medio de transmisión virtual de una trama de longitud t_I.
Tiempo medio de transmisión correcta de una trama:

$$t_V = t_I + E[T_1]p + \frac{T_2 p^2}{1-p}$$

Modelado de N.Red y Control de Flujo

Las redes se diseñan con recursos suficientes para soportar un tráfico nominal. Se pueden producir situaciones de congestión y bloqueo.

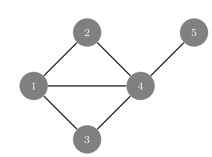


Niveles en el control de flujo

CFL: Control de Flujo Local
CFAR: Control de Flujo de Acceso a la Red
CFEE: Control de Flujo Entre Extremos
CFNO/ND: Control de Flujo entre Nodos Origen y Destino

$$\text{Potencia} = \frac{\text{Caudal}^{\alpha}}{\text{Retardo}}$$

Modelo del Circuito Virtual



Modelo de ventana deslizante es un modelo de control según actúa el transmisor al encontrarse con la ventana saturada. Se supone el caso más sencillo, el bloqueo de transmisión.
Hipótesis:
Los `ACK`s del CV asociados al control de ventana se supone que se transmiten de vuelta por la red con prioridad máxima.
Se desprecia el retardo de transmisión de `ACK`s.
Se modela el control de ventana deslizante sobre un CV como un **sistema cerrado**.
IMAGEN CV VENTANA
La fuente y destino están unidas por una cola artificial (Cola $M+1$) con una tasa de servicio λ que coincide con la de llegadas al CV. En situación de máxima sobrecarga hay un número fijo de paquetes N circulando por el sistema cerrado. Si hay N paquetes en el CV estarán en las M colas del sistema, la cola $M+1$ estará vacía (condición de bloqueo). Cuando un paquete alcanza al receptor pasará a la cola $M+1$ ($t_s = 0$) y ésta podrá generar servicios. Si hay menos de N paquetes en la ruta, el resto estarán en la cola $M+1$.
Teorema de Norton de agregación o descomposición de redes de colas, en redes con solución en forma de producto puede reemplazarse una subred por una cola compuesta con tasa de servicio dependiente del estado. La red mantiene el mismo comportamiento estadístico extremo a extremo. Da
Ecuaciones de estado de la cola del CV:

$$p_n = \frac{\lambda_e^n}{\prod_{i=1}^{n} u(i)} \cdot p_0 \quad \text{, Condición de normalización} \sum_{n=0}^{N} p_n = 1$$

Hipótesis validas para su uso: Enlaces con igual capacidad e igual tráfico total.
Para redes con enlaces de capacidades diferentes (μ_1 y μ_2) siendo $\mu_1 \gg \mu_2$ el cuello de botella estará en los enlaces de menor capacidad μ_2, en los de mayor capacidad μ_1 no se producen esperas $\rightarrow$ pueden ser eliminados.
Análisis de un CV homogéneo

$$u(n, M) = \frac{n}{n + (M-1)} \cdot \mu \text{Dependiente del estado}$$

Demostraciones

Esperanza y varianza en eventos Poissonianos

$$E(k) = \sum_{k=0}^{\infty} k \cdot p(k) = \sum_{k=0}^{\infty} k \cdot \frac{(\lambda \cdot T)^k}{k!} \cdot e^{-\lambda T} =$$
$$= e^{-\lambda T} \cdot \left(\sum_{k=1}^{\infty} \lambda T \cdot \frac{k \cdot (\lambda \cdot T)^{k-1}}{k \cdot (k-1)!} \right) = e^{-\lambda T} \cdot \lambda \cdot T \cdot e^{\lambda T} =$$
$$= \lambda \cdot T$$

$$\sigma_k^2 = E\left[(k - E[k])^2\right] = E[k^2] - E^2[k] = \sum_{k=0}^{\infty} k^2 \cdot p(k) - (\lambda \cdot T)^2 =$$
$$= \sum_{k=0}^{\infty} k^2 \cdot \frac{(\lambda \cdot T)^k}{k!} e^{-\lambda T} - \lambda^2 T^2 =$$
$$= e^{-\lambda T} \cdot \left(\sum_{k=1}^{\infty} \lambda T \cdot \frac{k \cdot (\lambda \cdot T)^{k-1}}{k \cdot (k-1)!} \right) - \lambda^2 T^2 =$$
$$= \lambda T e^{-\lambda T} \cdot \left(\sum_{k=1}^{\infty} \frac{(k-1+1) \cdot (\lambda \cdot T)^{k-1}}{(k-1)!} \right) - \lambda^2 T^2 =$$
$$= \lambda T e^{-\lambda T} \cdot \left(\sum_{k=1}^{\infty} \frac{(k-1) \cdot (\lambda \cdot T)^{k-1}}{(k-1)(k-2)!} + \sum_{k=1}^{\infty} \frac{(\lambda T)^{k-1}}{(k-1)!} \right) - \lambda^2 T^2 =$$
$$= \lambda T e^{-\lambda T} \cdot \left(\lambda T \cdot \sum_{k=1}^{\infty} \frac{(\lambda \cdot T)^{k-2}}{(k-2)!} + e^{\lambda T} \right) - \lambda^2 T^2 =$$
$$= \lambda T e^{-\lambda T} (\lambda T e^{\lambda T} + e^{\lambda T}) - \lambda^2 T^2 =$$
$$= \lambda \cdot T$$

Demostración de la distribución de Poisson

Stop & Wait Tiempo medio para una transmisión correcta

$$t_v = t_T(1-p) + 2t_T(1-p)p + 3t_T(1-p)p^2 + \ldots =$$
$$= t_T(1-p) \sum_{i=1}^{\infty} i \cdot p^{i-1} = t_T(1-p) \frac{1}{(1-p)^2} =$$
$$= \frac{t_T}{1-p}$$

Go Back N Tiempo medio para una transmisión correcta

$$t_V = (1-p) \cdot t_I + (1-p) \cdot p(t_T + t_I) + (1-p) \cdot p^2(2t_T + \cdot t_I) + \ldots =$$
$$= (1-p) \cdot t_I + (1-p) t_T \sum_{i=1}^{\infty} i \cdot p^i + (1-p) \cdot t_I \sum_{i=1}^{\infty} p^i =$$
$$= (1-p) \cdot t_I + a \cdot t_I \cdot (1-p) \frac{p}{(1-p)^2} + t_I \cdot (1-p) \cdot \frac{p}{1-p} =$$
$$= t_I\left(1 - p + \frac{ap}{1-p} + p\right) = t_I \frac{1-p+ap}{1-p} =$$
$$= t_I \frac{1+(a-1)p}{1-p}$$