Introduction à la simulation

Thi-Mai-Trang Nguyen

Université Pierre et Marie Curie (UPMC)

20 janvier 2017

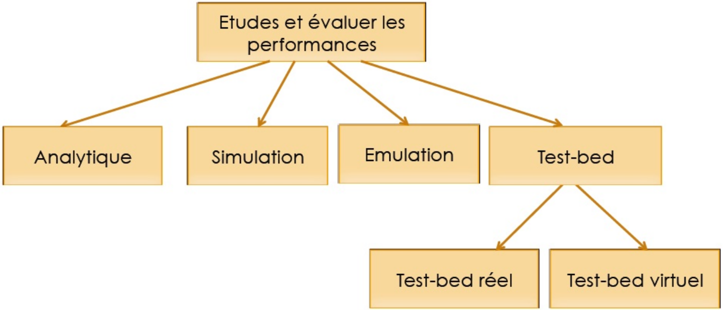
Plan

- Introduction
- Simulation
- Variables aléatoires
- Génération des nombres pseudo-aléatoires
- Moyenne, variance et écart type

Simulation - Emulation - Virtualisation des réseaux (SEV)

- Simulation, émulation et virtualisation sont les trois techniques couramment utilisés pour étudier les réseaux
  - Evaluation de performances
  - Test de fonctionnement
  - Validation de protocoles
  - Dimensionnement des réseaux
- Simulation : le comportement du réseau est modélisé par un logiciel (i.e. un simulateur)
- Emulation : utiliser un logiciel ou matériel qui peut introduire les comportements souhaités sous notre contrôle dans un réseau de test réel
- Virtualisation : monter un réseau expérimental en utilisant les machines virtuelles (i.e. un test-bed virtuel)

Méthodes pour études et évaluation de performances des réseaux

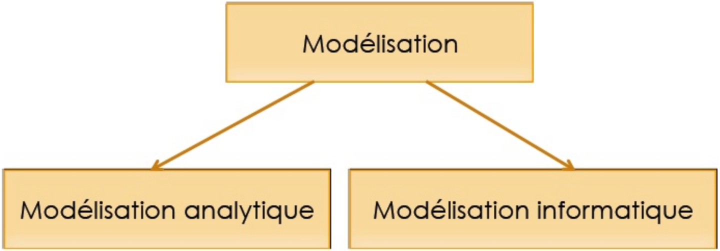


Test-bed et modélisation

- Test-bed
  - Travailler sur un prototype ou un réseau expérimental
  - Avantage : les résultats sont fidèles à la réalité
  - Inconvénient : coûteuse, expériences non reproductibles
- Modélisation
  - Utiliser les outils mathématiques ou les programmes informatiques pour modéliser le système
  - Avantage : moins coûteuse, utile à défaut de pouvoir disposer du système réel, paramétrable, reproductible
  - Inconvénient : une simplification plus ou moins importante par rapport au système réel

Modélisation

- Utiliser les modèles mathématiques (approche analytique) ou informatiques (approche simulation) pour construire une abstraction du système réel en négligeant certains détails jugés peu importants



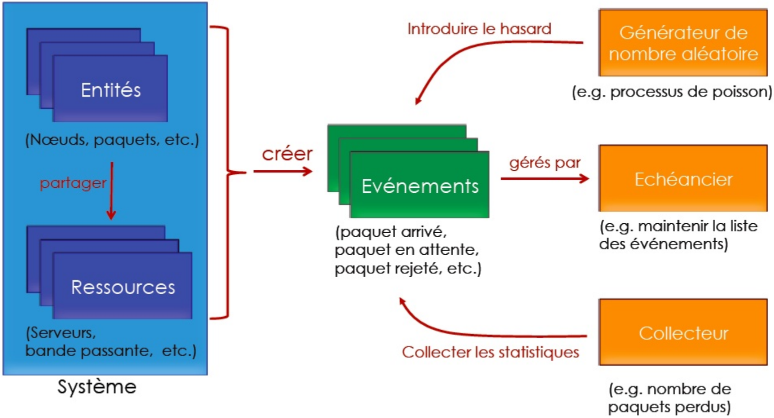
Modèles analytiques vs. Simulation

- Modélisation analytique
  - Etablir les modèles mathématiques qui représentent le système réel en utilisant les outils mathématiques comme la théorie des files d'attente ou la théorie de probabilité
  - Avantage : Résultats fermes, généraux et rapides à obtenir
  - Inconvénient : Simplification importante par rapport à la réalité
- Modélisation informatique (Simulation)
  - Essayer de reproduire le système réel en une version virtuelle du système par des programmes informatiques
  - Avantage : Pouvoir modéliser de manière très détaillé le système cible
  - Inconvénient : Il faut mener un certain nombre de campagnes de simulation qui sont souvent longues pour tirer une conclusion avec un minimum garanti de qualité statistique

Simulation

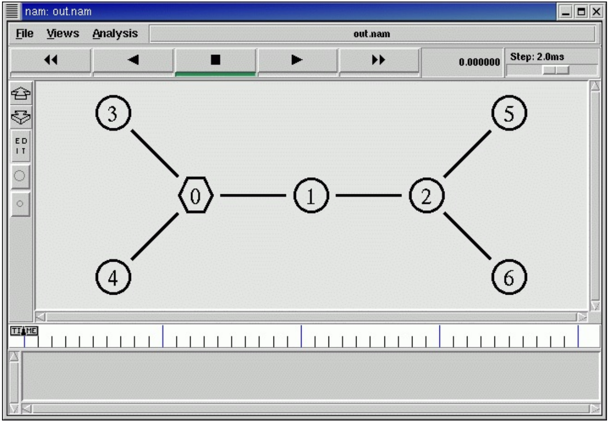
- Simulation est une approche pour modéliser et étudier un système réel par des modèles informatiques
- Le modèle du système est d'abord construit et ensuite implémenté sous la forme des programmes informatiques appelés un simulateur
- Les expériences réalisées par un simulateur permettent de comprendre le comportement d'un système et d'évaluer les performances du système

Structure d'une simulation



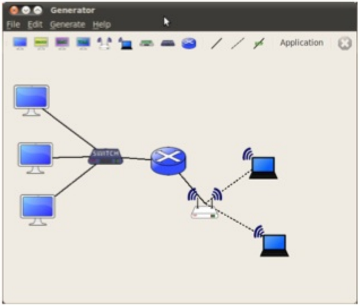
Les simulateurs réseaux les plus utilisés (1)

- NS-2
  - http ://www.isi.edu/nsnam/ns/



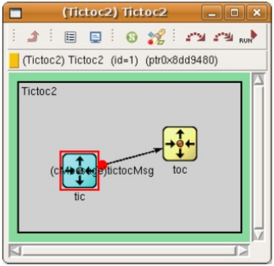
Les simulateurs réseaux les plus utilisés (2)

- NS-3
  - http ://www.nsnam.org/

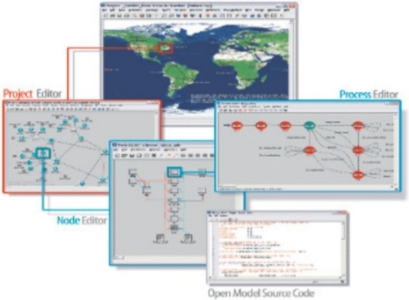


Les simulateurs réseaux les plus utilisés (3)

- OMNeT++
  - http ://www.omnetpp.org/



- OPNET
  - http ://www.riverbed.com/products/performance-management-control/opnet.html



Variable aléatoire dans simulation des réseaux

- Dans les simulations réseaux, plusieurs événements sont modélisés par les variables aléatoires
  - Arrivée d'un paquet
  - Temps d'attente
  - Temps de service
- Aléatoire vs. Déterministe
  - Une variable est **déterministe** quand on dispose d'une règle de calcul
  - Une variable est **aléatoire** quand sa valeur est peu prédictive et que les modèles déterministes ne sont pas efficaces pour calculer

Loi d'une variable aléatoire

- Donner la loi d'une variable aléatoire discrète X, c'est calculer les probabilités $P(X = x)$ pour toutes les valeurs x possibles prises par X (autrement dit, pour tous les x appartenant au support de X)
- Dans la réalité, beaucoup de phénomènes peuvent être efficacement modélisés à travers des lois probabilistes
    - Le nombre d'appels téléphoniques
    - Le nombre de clients consultant un site Web sur une période donnée
    - La taille des paquets dans l'Internet
    - La taille des fichiers
    - La durée d'une session

Comment caractériser une variable aléatoire

- Une variable aléatoire suivant une loi probabiliste est caractérisée par les paramètres suivants
  - La densité de probabilité (i.e. la fonction de masse) = PDF (Probability Density Function)
  - Fonction de répartition = PCF (Probability Cumulative Function)
  - Moyenne
  - Variance
  - Ecart type
- Quelques lois usuelles
  - Loi uniforme
  - Loi exponentielle
  - Loi de Pareto
  - Loi de Poisson
- Identification de loi est une technique relevant de la statistique

Génération de nombres pseudo-aléatoires

- Pour alimenter une simulation, il est nécessaire de générer une suite de variables aléatoires pour reproduire artificiellement la loi identifiée
- Avant l'arrivée des ordinateurs : utilisation des suites pré-établies des valeurs obtenues des expériences réelles
  - Coûteuse, non reproductible
  - Souvent, les suites ne sont pas suffisamment longues
- Avec les ordinateurs : utilisation d'une génération algorithmique des nombres aléatoires
  - Facile, programmable, reproductible
  - Cependant, la série générée se répète au bout d'un certain cycle plus ou moins long
  - Ces nombres sont qualifiés comme des nombres pseudo-aléatoires
  - Le générateur est appelé "générateur des nombres pseudo-aléatoires" (PRNG – Pseudo Random Number Generator)

PRNG à base des variables aléatoires uniformes sur (0,1)

- U(0,1) est très importante !
- Les variables aléatoires suivant une loi quelconque peuvent être générées à partir de celles de U(0,1)
- Les générateurs des variables aléatoires
  - Von Neumann (un des premiers générateurs)
  - LCG (OMNeT++)
  - Marsenne Twister (PRNG par défaut dans OMNeT++)
  - Ecuyer (OMNeT++, NS-3)

Loi uniforme

- On dit que la variable aléatoire X suit la loi uniforme sur son support $\{x_1, x_2, ..., x_n\}$ lorsque sa densité de probabilité est $P(X = x_i) = \frac{1}{n}$ pour tout $1 \leq i \leq n$
- Exemple

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

TABLE: X est le résultat du lancer d'un dé

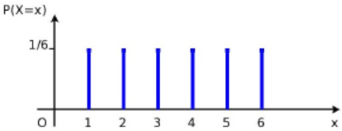


FIGURE: Densité de probabilité de X

Générateur de Von Neumann (1)

- Proposé en 1940s par Von Neumann
- C'est un des premiers générateurs réalisables par l'ordinateur
- Algorithme
  - Commencer avec un entier n_0 (encore appelé un "seed", un "germe" ou une "graine") possédant 4 digits ($0 < n_0 < 10000$)
  - n_1 sera les 4 digits du milieu de ${n_0}^2$ et ainsi de suite pour obtenir une suite n_i
  - En cas d'un nombre impaire de digits, on retire le digit le plus à droite
  - A partir de la suite n_i, on obtient la suite $r_i = \frac{n_i}{10000}$
  - r_i est une suite de nombres aléatoires uniformes sur $(0, 1)$

Générateur de Von Neumann (2)

- Exemple

i	n_i	r_i	${n_i}^2$
0	4321	0.4321	18**6710**41
1	6710	0.6710	45**0241**00
2	241	0.0241	**5808**1
3	5808	0.5808	33**7328**64
4	7328	0.7328	53**6995**84

- Remarques
  - Toute séquence $\{n_i\}$ est une suite de nombres déterministes fixée par le seed n_0
  - Toute séquence $\{n_i\}$ est périodique avec une période P. C-a-d. pour tout i, la même séquence $\{n_i, ..., n_{i+P-1}\}$ se reproduira à partir de n_{i+P}. Dans l'exemple ci-dessus, $P < 10000$.

Les générateurs LCG

- Les générateurs congruentiels linéaires (LCG – Linear Congruantial Generator) ont été introduits en 1948 par Derrick Lehmer
- La forme générique :

$$X_{n+1} = (aX_n + b) \mod M$$

- Trois paramètres : a (le multiplicateur), b (l'incrément) et M (le module).
- Comme les valeurs $\{Xn\}$ sont inférieures à M, la série des variables aléatoires sur $[0, 1)$, $\{r_n\}$, est obtenue par

$$r_n = \frac{X_n}{M}$$

- X_0 est le "seed"

Propriétés des générateurs LCG

- Un générateur LCG est entièrement caractérisé par (a, b, M)
- Pour chaque (a, b, M), la suite de nombres $\{X_i\}$ générée est complètement déterminée par X_0
- La période P $(P < M)$ atteint sa valeur maximale si les conditions suivantes sont satisfaites
  - b et M sont premiers entre eux (leur seul diviseur commun est 1)
  - tout diviseur premier de M l'est également pour $a - 1$
  - si M est un multiple de 4, alors $(a - 1)$ l'est aussi

Les générateurs MLCG

- Lorsque $b = 0$, on parle de LCG multiplicatif(MLCG - Multiplicative LCG) qui sont plus facile à réaliser dans les ordinateurs

$$X_{n+1} = aX_n \mod M$$

- Deux types de MLCG
  - MLCG avec $M = 2^k$ (très facile à réaliser par les ordinateurs mais la période maximale est limitée à $P = \frac{M}{4}$)
  - MLCG avec M nombre premier (pouvoir atteindre la période maximale de $M - 1$)

Génération d'une variable aléatoire discrète suivant une loi quelconque à partir de sa fonction de masse

- Pour générer les valeurs d'une variable aléatoire discrète X caractérisée par sa densité de probabilité

$$P\{X = x_j\} = p_j, \qquad j = 0, 1, \ldots, \qquad \sum_j p_j = 1$$

nous devons tout d'abord générer une variable aléatoire $u \sim U(0, 1)$

- Les valeurs de X sera ensuite déterminées en fonction des valeurs de u de la manière suivante

$$X = \begin{cases} x_0 & \text{if} \qquad u < p_0 \\ x_1 & \text{if} \qquad p_0 \leq u < p_0 + p_1 \\ \ldots & \\ x_j & \text{if} \qquad \sum_{i=1}^{j-1} p_i \leq u < \sum_{i=1}^{j} p_i \\ \ldots & \end{cases}$$

Loi Uniforme sur un intervalle U(a,b) (1)

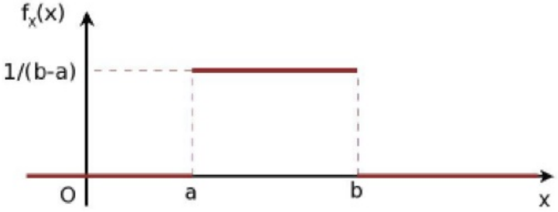
- Le modèle le plus simple pour modéliser un phénomène aléatoire continu ayant des valeurs qui varient entre a et b
- Utilisée lorsqu'on a peu de connaissance précise sur le comportement statistique d'un phénomène aléatoire
- Une variable aléatoire X est uniformément répartie sur l'intervalle $[a, b]$, $a < b$, si sa densité de probabilité $f_X(x)$ est telle que

$$f_X(x) = P(X = x) = \begin{cases} \frac{1}{b-a} & \text{pour} x \in [a, b] \\ 0 & \text{ailleurs} \end{cases}$$

- a et b sont les deux paramètres qui correspondent aux deux extrémités de la plage de variation de la grandeur physique modélisée

Loi Uniforme sur un intervalle U(a,b) (2)

- La densité de probabilité de U(a,b)



- Génération de $X \sim U(a,b)$ à partir de $u \sim U(0,1)$

$$X = a + (b-a)u$$

Loi de Poisson (1)

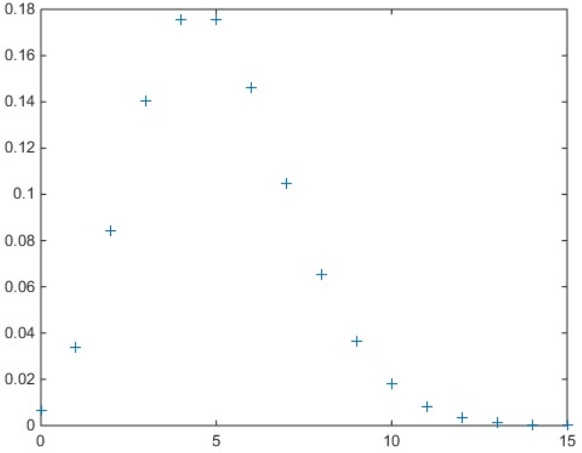
- Très efficace pour modéliser
  - Le nombre d'appels téléphoniques
  - Le nombre de clients consultant un site Web sur une période donnée
- Une variable aléatoire X prenant valeurs dans $\mathbb{N}$ suit une loi de Poisson de paramètre λ ($\lambda > 0$), notée $P(\lambda)$, si sa densité de probabilité est telle que

$$P(X = k) = e^{-k} \frac{\lambda^k}{k!}$$

- λ est la valeur moyenne de la variable

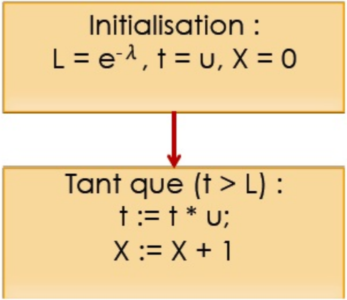
Loi de Poisson (2)

- Un exemple de densité de probabilité de $P(5)$



Loi de Poisson (3)

- Génération de $X \sim P(\lambda)$ à partir de $u \sim U(0, 1)$



Loi exponentielle (1)

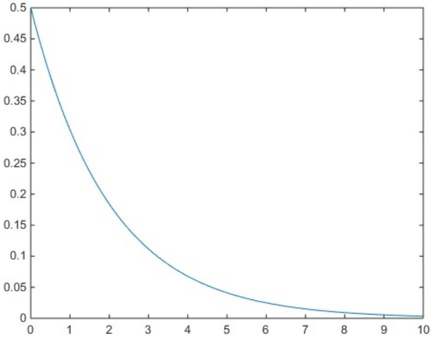
- Très efficace pour modéliser les intervalles de temps d'un processus d'arrivées de Poisson
- Une variable aléatoire X, $x > 0$, suit la loi exponentielle $Exp(\lambda)$ si sa densité de probabilité est telle que

$$P(X = x) = \lambda e^{-\lambda x}$$

- Le paramètre λ ($\lambda > 0$) correspond à l'inverse de la moyenne de la variable

Loi exponentielle (2)

- Un exemple de la densité de probabilité de $Exp(2)$



- Génération de $X \sim Exp(\lambda)$ à partir de $u \sim U(0,1)$

$$X = -\frac{\log u}{\lambda}$$

Loi de Pareto (1)

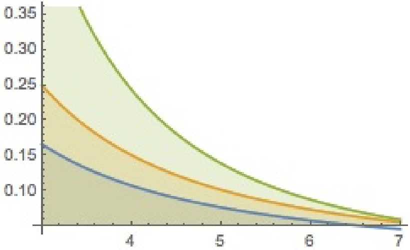
- Utilisée pour modéliser les phénomènes ayant la probabilité de "longue queue"
  - Les grandes valeurs possèdent une part non négligeable parmi toutes les valeurs possibles (e.g. la taille des messages dans l'Internet)
- Une variable aléatoire suit la loi de $Pareto(s, \alpha)$ si sa densité de probabilité est telle que

$$P(X = x) = \frac{\alpha}{s}(\frac{x}{s})^{-(\alpha+1)} = \alpha \frac{s^{\alpha}}{x^{\alpha+1}}, x \geq s$$

- Deux paramètres :
  - s est le seuil au dessus duquel la loi est définie
  - α est l'indice de Pareto qui détermine la forme

Loi de Pareto (2)

- Un exemple des densités de probabilité de $Pareto(3, 0.5)$, $Pareto(3, 0.75)$ et $Pareto(3, 1.5)$



- Génération de $X \sim$ Pareto (s, α) à partir de $u \sim U(0, 1)$

$$X = -\frac{s}{u^{\frac{1}{\alpha}}}$$

Paramètres importants caractérisant une variable aléatoire

- Moyenne
  - La valeur moyenne des valeurs observées

$$E(X) = \overline{X} = \frac{\sum_1^n X_i}{N}$$

- Variance
  - Représente l'écart quadratique des valeurs d'une variable aléatoire par rapport à sa moyenne

$$Var(X) = \sigma^2 = \frac{\sum_1^n (X_i - \overline{X})^2}{N}$$

- Ecart type
  - Représente l'écart des valeurs d'une variable aléatoire par rapport à sa moyenne

$$\sigma = \sqrt{Var(X)}$$

Moyennes et variances de quelques lois usuelles

- Loi uniforme $U(a, b)$
  $E[X] = \frac{a+b}{2} \quad Var[X] = \frac{(b-a)^2}{12}$
- Loi de Poisson $P(\lambda)$
  $E[X] = \lambda \quad Var[X] = \lambda$
- Loi exponentielle $Exp(\lambda)$
  $E[X] = \frac{1}{\lambda} \quad Var[X] = \frac{1}{\lambda^2}$
- Loi de Pareto $Pareto(s, \alpha)$

$$E(X) = \begin{cases} \frac{\alpha s}{\alpha - 1} & \text{pour} \quad \alpha > 1 \\ \infty & \text{pour} \quad \alpha \leq 1 \end{cases}$$

$$Var(X) = \begin{cases} \frac{\alpha s^2}{(\alpha-1)^2(\alpha-2)} & \text{pour} \quad \alpha > 2 \\ \infty & \text{pour} \quad \alpha \leq 2 \end{cases}$$

Références

- Ken Chen, "Evaluation de performances par simulation et analyse", ISTE Editions, Hermes Science, 2014
- Sheldon M. Ross, "A course in simulation", Macmillan, 1990
- Teerawat Issariyakul et Ekram Hossain, "Introduction to Network Simulator NS2", Springer, 2009