DÉTECTION D'INTRUSIONS PAR ALGORITHME GÉNÉTIQUE

Ludovic Mé

Supélec
Service Informatique
Rennes

mars 1996

Pour une détection efficace

La détection est faite en exploitant le contenu des fichiers d'audit

<u>Le problème</u>
énorme volume de données dans le fichier d'audit

<u>Les conséquences</u>

- des outils
- des méthodes $\Longrightarrow$ approche comportementale
  $\Longrightarrow$ approche par scénarios

Méthodes de détection d'intrusions

- Détection d'une attaque exploitant une vulnérabilité inconnue : approche comportementale

  $\Downarrow$

  – proposée par J.P. Anderson (80) puis D.E. Denning (87)
  – basée sur l'hypothèse selon laquelle l'exploitation d'une vulnérabilité du système implique un usage anormal de celui-ci.

- Détection d'une attaque exploitant une vulnérabilité connue : approche par scénarios

  $\Downarrow$

  le comportement de l'utilisateur correspond-il à une attaque connue ?

Avantages de chacune des deux approches

- Approche comportementale

  $\Downarrow$

  détection d'intrusion inconnue : possible

- Approche par scénarios

  $\Downarrow$

  prise en compte des comportements exacts des attaquants potentiels : possible

Approche comportementale : inconvénients

- Choix des différents paramètres du modèle statistique : délicat
- Choix des mesures à retenir pour un système cible donné : délicat
- Si profonde modification de l'environnement du système cible : flot ininterrompu d'alarmes venant du modèle statistique
- Un utilisateur peut changer lentement de comportement dans le but d'habituer le système à un comportement intrusif

Approche par scénarios : inconvénients

- La base de règle doit être bien construite : délicat
- Performances du système expert : limitées par celles de l'expert humain qui fourni les règles

Outil	Méthode statistique	Système expert	Réseau de neurones	Algorithme génétique
IDES	X	X		
AudES		X		
Computer Watch		X		
DIDS	X	X		
HAYSTACK	X	X		
ISOA	X	X		
MIDAS	X			
NADIR	X	X		
NIDX	X	X		
W&S	X			
Hyperview	X	X	X	
GASSATA				X

GASSATA : première formalisation du problème

- scénario d'attaque $\Longleftrightarrow$ expression régulière (rewbr)
  modèle d'attaque
- les attaques peuvent être entremêlées $\Longleftrightarrow$ métamodèle d'attaque
  ab et cd $\Rightarrow$ ab, cd, abcd, acbd, acdb, cdab, cadb, cabd
- recherche de scénarios dans le fichier d'audit
  $\Downarrow$
  recherche multiple de métamodèles dans une chaîne de caractères
- explosion combinatoire du nombre de cas à traiter
  NP-Complet

Nombre de cas à traiter

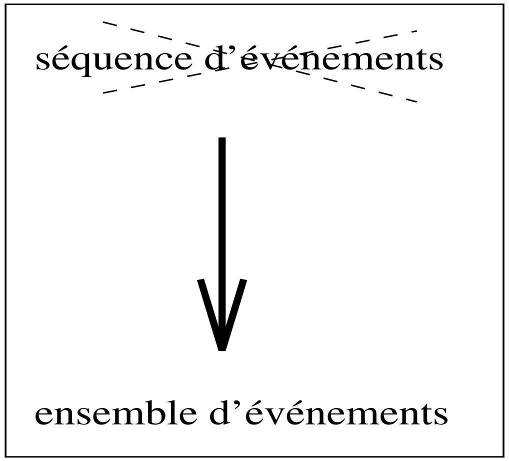
N_i	$N=2$	$N=4$	$N=6$	$N=8$	$N=10$
1	4	64	1956	109600	$9,8641.10^{06}$
2	8	2920	$8,2045.10^{06}$	$8,73944.10^{10}$	$2,50469.10^{15}$
4	72	$6,3202.10^{07}$	$3,2485.10^{15}$	$2,39099.10^{24}$	$1,287.10^{34}$
6	926	$2,30881.10^{12}$	$2,67019.10^{24}$	$1,71889.10^{38}$	$2,22254.10^{53}$
8	12872	$9,95611.10^{16}$	$2,88925.10^{33}$	$1,81657.10^{52}$	$6,30264.10^{72}$
10	184758	$4,70536.10^{21}$	$3,64415.10^{42}$	$2,38025.10^{66}$	$2,35707.10^{92}$
12	$2,70416.10^{06}$	$2,35809.10^{26}$	$5,0696.10^{51}$	$3,57829.10^{80}$	$1,05202.10^{112}$
14	$4,01166.10^{07}$	$1,23094.10^{31}$	$7,54977.10^{60}$	$5,91823.10^{94}$	$5,30917.10^{131}$
16	$6,0108.10^{08}$	$6,62123.10^{35}$	$1,1821.10^{70}$	$1,05003.10^{109}$	$2,93265.10^{151}$

$$N_{mm} = \Sigma_{K=1}^{N}(\underbrace{V_K^N + V_K^N + \ \ldots \ + V_K^N}_{C_N^K termes}) \quad \text{avec} \quad V_K^N = \frac{\left(\Sigma_{i=1}^{K} N_i\right)!}{\Pi_{i=1}^{K}(N_i!)}$$

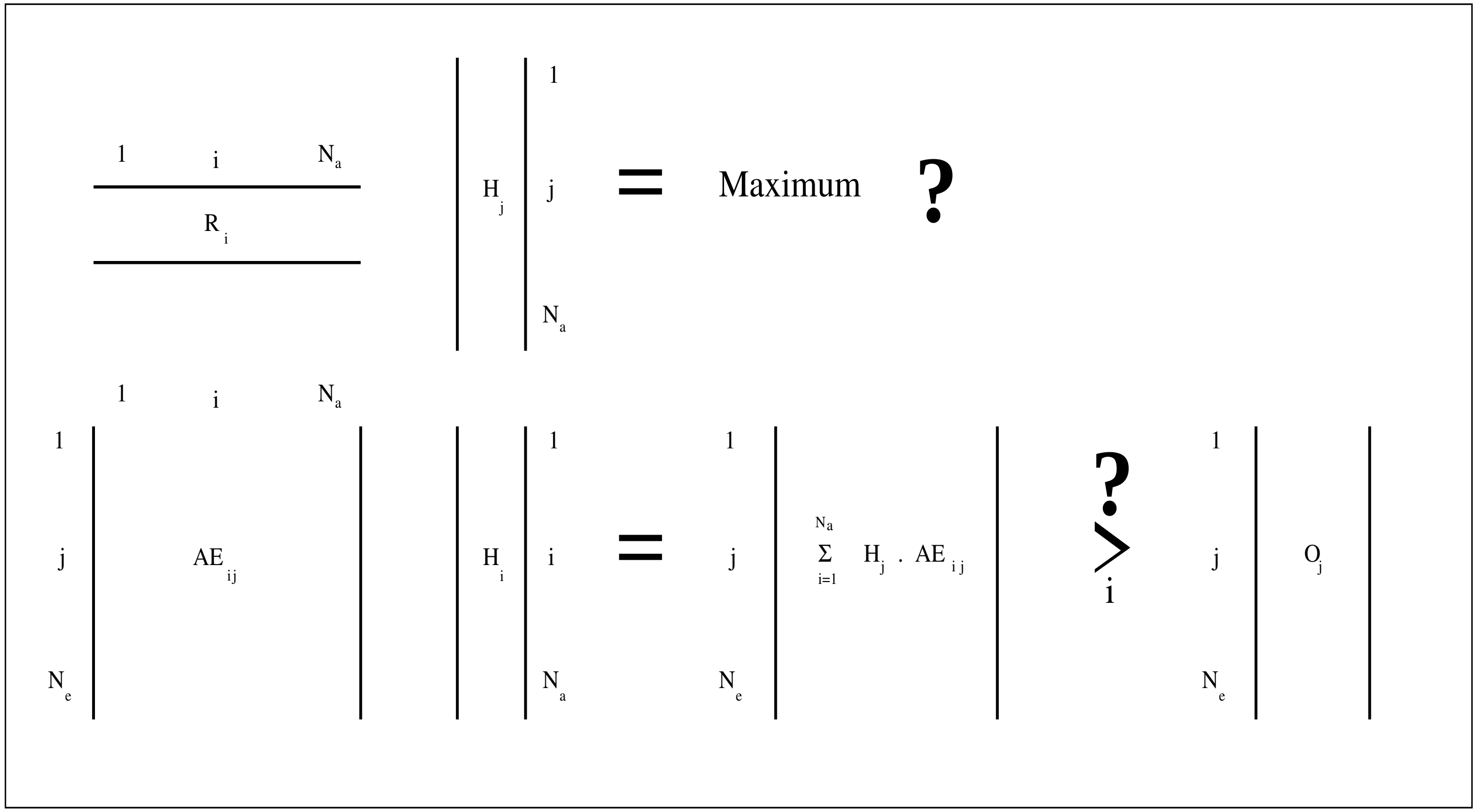
Conséquences

- application des algorithmes classiques de recherche de sous-suites impossible sur machine séquentielle

- une simplification possible du problème par analogie avec la stratégie de diagnostic d'un expert humain



GASSATA : Analyse simplifiée d'un fichier d'audit



Un exemple

$$
\begin{array}{ccccc} 1 & & & & N_a \\ \hline 1 & 1 & 1 & 1 & 1 \\ \hline \end{array}
\quad
\begin{array}{c|c|l} & 1 & 1 \\ & 0 & \\ & 1 & \\ & 0 & \\ & 0 & N_a \end{array}
= 2
$$

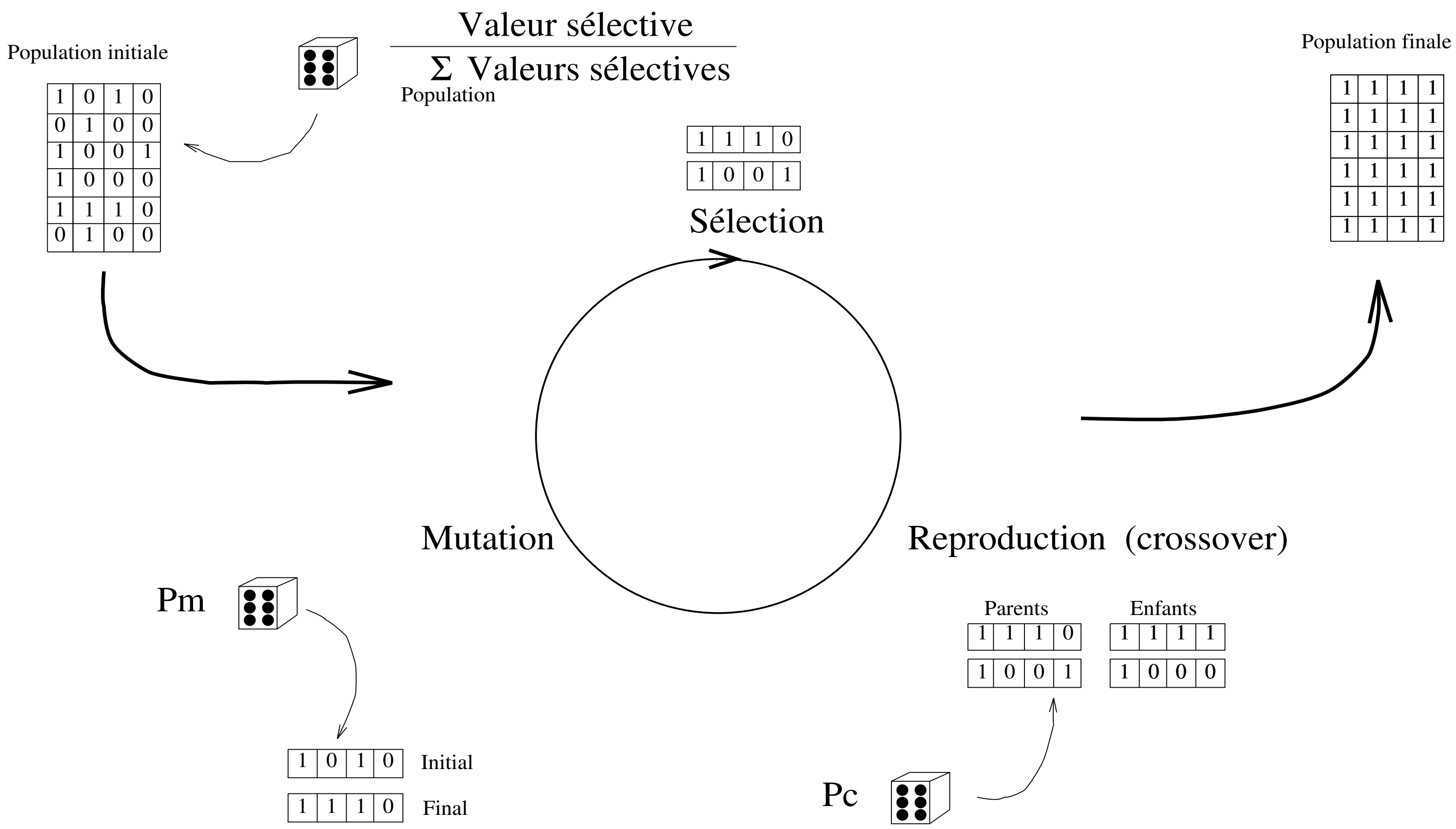
$$
\begin{array}{c|ccccc|}
 & 1 & & i & & N_a \\
1 & 1 & 0 & 0 & 1 & 0 \\
 & 0 & 0 & 10 & 0 & 0 \\
 & 0 & 3 & 0 & 0 & 0 \\
 & 5 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 1 & 0 & 0 \\
 & 0 & 7 & 0 & 0 & 0 \\
 & 9 & 0 & 2 & 1 & 0 \\
 & 0 & 0 & 0 & 30 & 0 \\
N_e & 0 & 0 & 0 & 0 & 1
\end{array}
\quad
\begin{array}{|c|l} 1 & 1 \\ 0 & \\ 1 & \\ 0 & \\ 0 & N_a \end{array}
=
\begin{array}{c|c|} 1 & 1 \\ & 10 \\ & 0 \\ & 5 \\ & 1 \\ & 0 \\ & 11 \\ & 0 \\ N_e & 0 \end{array}
\ \underset{i}{\overset{?}{>}}\
\begin{array}{c|c|} 1 & 2 \\ & 0 \\ & 0 \\ & 15 \\ & 1 \\ & 0 \\ & 11 \\ & 0 \\ N_e & 1 \end{array}
$$

NP-complétude du problème simplifié

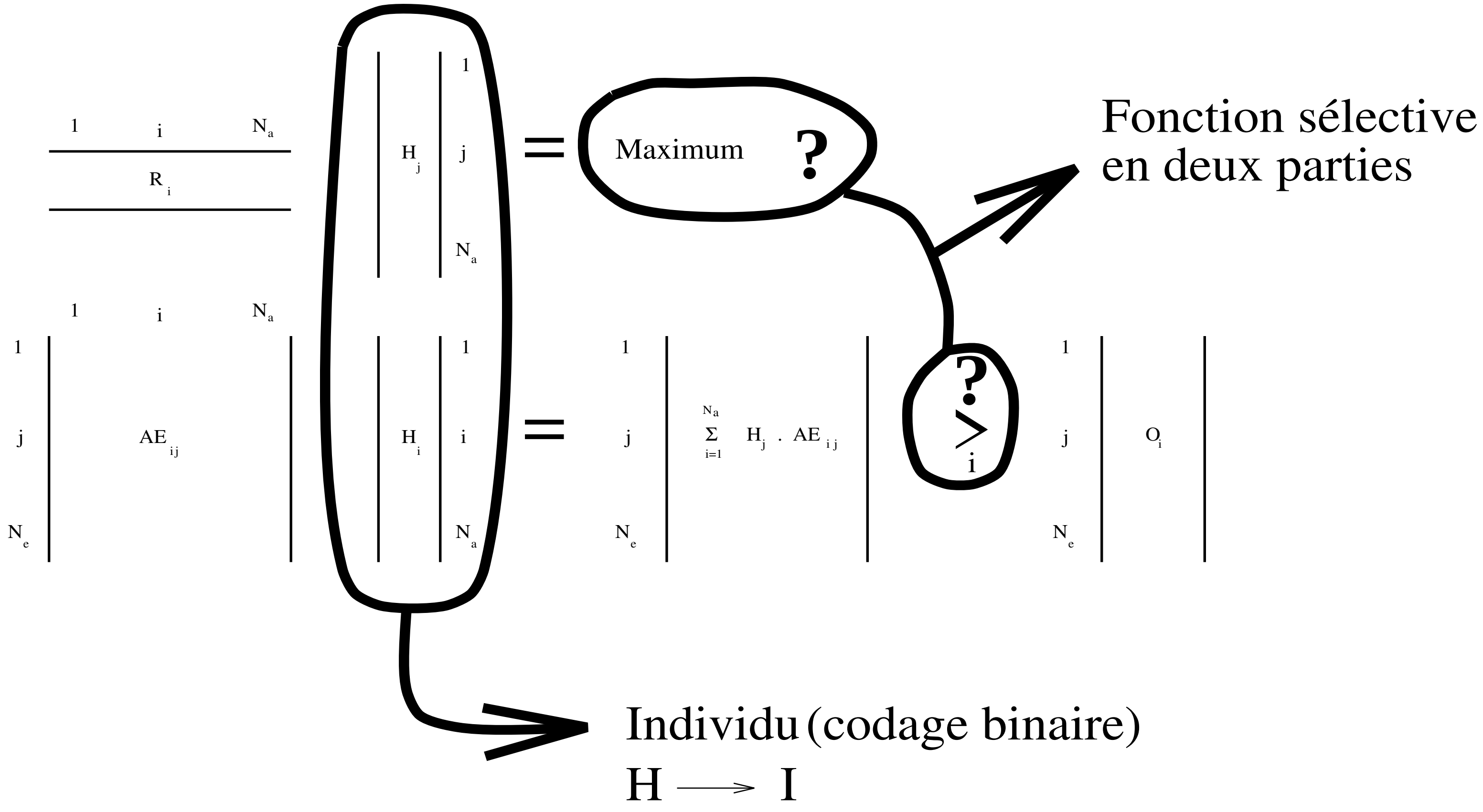
nombre d'attaques	temps de traitement
1	33.6 μsec.
5	3.12 msec.
10	0.19 sec.
15	8.8 sec.
20	6.16 minutes
24	1 heure 57 minutes
30	155 heures 20 minutes
40	24 années
50	305 siècles et 8 ans

50 événements par attaque
IBM RS6000 11.7 Mflops

Un algorithme génétique simple



Audit par algorithme génétique



Démarche utilisée pour une validation expérimentale

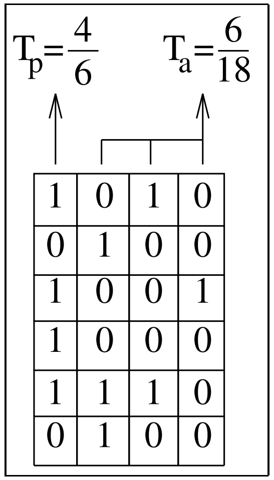
1. choix de la source des données d'audit
2. définition d'attaques et d'actions litigieuses
3. définition de comportements correspondant à des utilisateurs types
4. insertion d'attaque(s) dans les comportements types
5. analyse des fichiers d'audit résultant avec la fonction $F(I) = 200 + \left(\Sigma_{i=1}^{N_a} R_i.I_i \; - 2.T_e{}^2\right)$

Une matrice Attaques/Evénements

User_Login fail	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
User_Login (23h à 6h)	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Short_Session	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
User_SU OK	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
User_SU fail	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
who, w, finger, f, ps. . .	.	3	.	.	.	.	.	.	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
more, pg, cat, vi, . . .	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	5	.	.
ls OK	.	.	.	.	.	30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ls fail	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
df, hostname, uname	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
arp, netstat, ping	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.
ypcat	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.
lpr	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	1	.	.	.	.	.	.	.	.
rm, mv	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
ln	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.
whoami, id	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4
rexec, rlogin, rsh, . . .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
Proc_Execute	.	3	.	.	.	35	5	.	8	3	2	3	.	.	10	3	.	300	.	2	.	5	.	4
Proc_SetPri	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.
File_Open fail	.	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
File_Open fail cp	.	.	.	.	.	.	.	.	.	.	.	.	.	10	.	.	.	.	.	.	.	.	.	.
File_Open .netrc	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
File_Read lpr	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	.	.	.	.	.	.	.	.	.
File_Read passwd, . . .	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.
File_Write passwd, . . . fail	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
File_Write cp OK	.	.	.	.	.	.	.	.	.	.	.	.	30	.	.	.	.	.	.	.	.	.	.	.
File_Unlink rm	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	50	.	.	.	.	.	.	.
File_Mode	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.

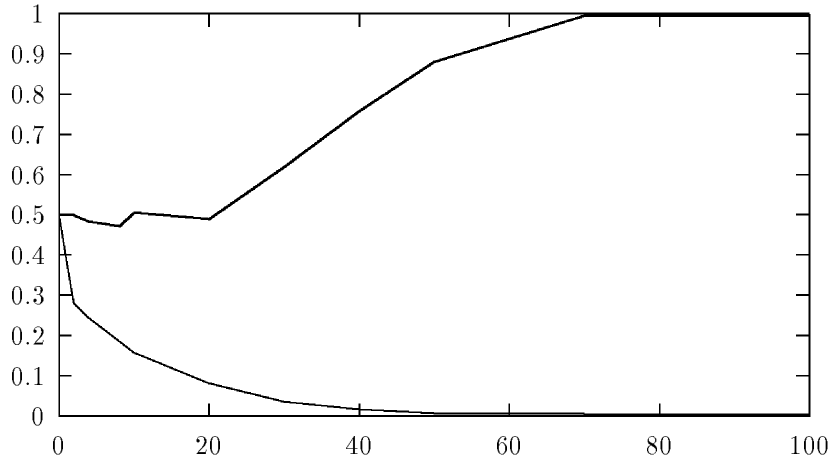
Evaluation des résultats

- T_p $\Rightarrow$ proportion de gènes d'allèle 1 pour les locus correspondant à une attaque présente
  T_a $\Rightarrow$ proportion de gènes d'allèle 1 pour les locus correspondant à une attaque absente



- évolution de la valeur sélective
- temps de traitement

Evolution des taux T_p et T_a



population de 500 individus
Probabilité de crossover = 0.6
Probabilité de mutation = 0.002
2 attaques présentes dans le fichier analysé

Temps de traitement

nombre d'attaques	temps de traitement	taux d'exploration
24	18"	3×10^{-3}
40	32"	5×10^{-8}
100	104"	5.9×10^{-26}
200	625"	6.3×10^{-56}

$P_c = 0.7$, $P_m = 0.002$, 500 individus

λ générations pour que résultat de qualité constante

28 types d'événements possibles

IBM RS6000 320

Pour plus d'informations

- Audit de sécurité par algorithmes génétiques. Thèse de doctorat, Université de Rennes 1, Numero d'ordre 1069, juillet 1994.
- Security Audit Trail Analysis Using Genethic Algorithms. In Proceedings of the 12th International Conference on Computer Safety, Reliability and Security, pp329-340, Poznan, October 1993.
- Détection d'intrusion dans un système informatique : méthodes et outils. TSI, 96(4), 1996, (à paraître, co-auteur Véronique Alanou).
- Ces publications sont accessibles depuis :

  `http://www.supelec-rennes.fr/rennes/si/equipe/lme/perso/publi.html`
- D'autres informations sont accessibles depuis :

  `http://www.supelec-rennes.fr/rennes/si/equipe/lme/these/these-lm.html`