Quelques projets d'ISPP

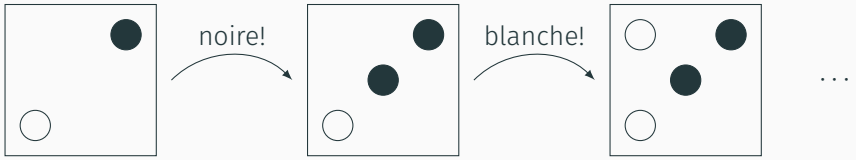
Rémi Géraud-Stewart <remi.geraud@ens.fr>

7 février 2019

Projet n°1 : Pólya et YouTube

Soit une urne contenant une boule noire et une boule blanche.

On tire une boule au hasard : si elle est noire (resp. blanche), on la remet et on ajoute une boule noire (resp. blanche) supplémentaire à l'urne. Et on recommence.

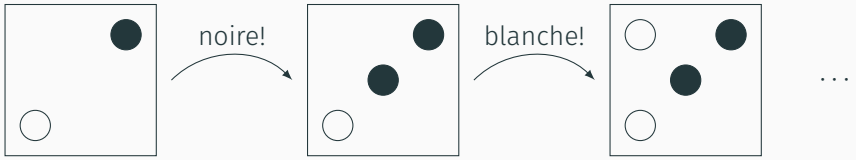


À la limite, quelle est la proportion ρ de boules noires sur le nombre de boules blanches ?

Projet n°1 : Pólya et YouTube

Soit une urne contenant une boule noire et une boule blanche.

On tire une boule au hasard : si elle est noire (resp. blanche), on la remet et on ajoute une boule noire (resp. blanche) supplémentaire à l'urne. Et on recommence.



À la limite, quelle est la proportion ρ de boules noires sur le nombre de boules blanches ?

Théorème (Pólya, 1930) : ρ est distribuée uniformément entre 0 et 1

Projet n°1 : Pólya et YouTube

Au lieu des balles de couleurs considérons des (catégories de) vidéos.

À chaque fois qu'une vidéo est choisie (selon une distribution donnée, e.g. uniforme), un certain nombre de vidéos est ajoutée à l'urne (selon une distribution donnée).

Par exemple, on peut ajouter $f(n)$ vidéos de la même catégorie que celle selectionnée, où f est une fonction et n est le nombre de fois que cette catégorie a été tirée dans l'histoire.

Question : converge-t-on vers certaines catégories (« bulle algorithmique »), ou bien, comme dans l'exemple de Pólya, explore-t-on en réalité tous les choix possibles uniformément ?

Projet n°2 : Signatures et isomorphismes de corps

Deux corps finis de même taille sont isomorphes

(e.g. $\mathbb{F}_2[X]/(X^7 + X + 1) \simeq \mathbb{F}_2[X]/(X^7 + X^2 + 1)$)

En revanche un tel isomorphisme n'est pas canonique

En 2017, Doroz et al. on proposé un système de chiffrement basé sur la difficulté supposée de calculer de tels isomorphismes[1].

Question 1 : peut-on construire un schéma de signature basé sur cette hypothèse de difficulté ?

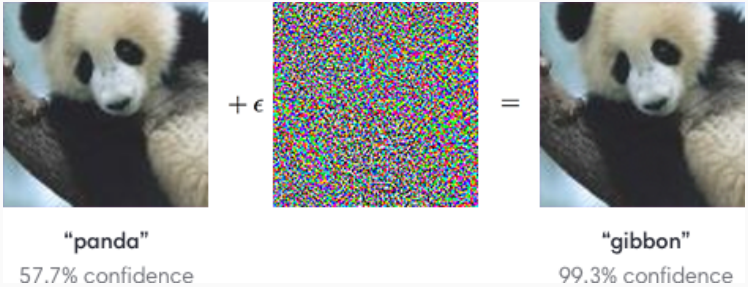
Question 2 : peut-on attaquer cette hypothèse en calculant efficacement les isomorphismes ?

1. `https://eprint.iacr.org/2017/548`

Projet n°3 : CAPTCHAs généralisés

Les systèmes de reconnaissance d'image modernes atteignent des performances quasi-humaines dans beaucoup de cas.

Néanmoins, ils sont parfois manipulables comme dans l'exemple suivant sur GoogLeNet (Szegedy et al. 2014)



Question : est-il possible de produire des media (images, videos, sons, etc.) aisément reconnaissables par des humains mais qui trompent systématiquement plusieurs systèmes automatisés ?

Projet n°4 : Obfuscation dynamique

L'obfuscation est une réécriture d'un programme visant à en dissimuler le mécanisme.

Opération généralement effectuée lors de la compilation, qui introduit à ce moment une grande complexité, des leurres, etc. mais une fois compilé (et obfusqué) le programme n'est plus modifié peut être analysé :(

Idée : transformer un programme P en un programme $Q(x)$ tel que

$$Q(x) = \begin{cases} P & \text{si } x \in X \\ P' & \text{sinon} \end{cases}$$

pour un langage X fixé, P' étant autre chose (e.g. un programme ne terminant pas)

(Et pour rigoler, on devrait pouvoir itérer la construction)

Projet n°5 : Comparaison de périodes

Une *période* est un nombre qui s'exprime comme une intégrale (absolument convergente) d'une fraction rationnelle à coefficients rationnels sur un domaine de $\mathbb{R}^n$ défini par des inégalités rationnelles à coefficients rationnels.

(e.g. $\sqrt{2} = \int_{2x^2 \leq 1} \mathrm{d}x$)

C'est donc un ensemble dénombrable, qui contient tous les entiers, tous les rationnels, mais aussi $\sqrt{2}$, π, $\log(7)$, $\zeta(3)$... (on ignore si *e* est une période)

Question : existe-t-il un algorithme prenant deux périodes et déterminant si elles sont égales ? quelle est sa complexité ? peut-on trouver des relations entre périodes (e.g. exprimer $\zeta(3)$ à partir de π etc.) ?

(David : on peut leur proposer aussi de conclure les crazy sums ?)

Projet n°6 : Géométrie des logs

En sécurité informatique, on a souvent à traiter des logs très complexes contenant beaucoup d'informations. Par exemple, un flux réseau contiendra des adresses IP, des extraits de trame, des échos etc.

On peut envisager de traiter ces données comme un nuage de points dans l'espace n-dimensionnel (souvent pour comparer deux à deux de tels nuages).

Question : est-ce que des invariants topologiques ou géométriques (persistence homologique, graphe de Voronoi, volume...) efficacement calculables permettent d'identifier des logs pertinents/suspects, caractériser un attaquant ou un type d'attaque ?