
afnic
Mesures DNS à l'ère du Big Data : outils et défis
JCSA, 9 juillet 2015
Vincent Levigneron, Afnic
afnic

Sommaire

1. Mesures DNS réalisées par l'Afnic
2. Volumes et biais
3. Limitations
4. Pourquoi une approche Big-Data
5. Réalisation d'un PoC Big-Data
6. Défis à relever/attentes
7. Perspectives et pistes

Mesures DNS réalisées par l'Afnic

- ✓ Des indicateurs de résilience pour l'Observatoire (ODRIF)
  - ✓ Dispersion topologique des serveurs DNS faisant autorité
  - ✓ Taux de pénétration d'IPv6
  - ✓ Taux d'adoption de DNSSEC
- ✓ Des sources d'informations
  - ✓ Base du registre .fr
  - ✓ Fichiers de zone
  - ✓ Traces du trafic sur les serveurs DNS du .fr faisant autorité
  - ✓ Plateforme DNSwitness (outils de mesures et base de stockage / interrogation)
  - ✓ Bases Whois externes (adresses IP, numéros d'AS)

afnic

Volumes et biais

- ✓ Base du registre et fichiers de zone
  - ✓ Près de 3 Millions de noms de domaine délégués
  - ✓ Possibilité d'utiliser 100% des informations
- ✓ Indicateurs basés sur le trafic DNS
  - ✓ 5 instances du serveur anycast d.nic.fr faisant autorité pour .fr
  - ✓ Utilisation de 5 sondes DNSmezzo : une sonde au niveau de chaque instance anycast (capture, analyse et insertion dans une base de données)
    - ✓ Existence d'un biais (assumé) dans les résultats, du fait de la localisation des instances
  - ✓ Les 8 instances du serveur d.nic.fr reçoivent en moyenne 4000 r/s

Limitations

- ✓ Nécessité d'échantillonner
  - ✓ Uniquement les mois nécessaires aux calculs
  - ✓ 5% sur 24 heures, 4 jours de récolte
  - ✓ Cela représente malgré tout 500 Go de données/an à conserver
- ✓ Les bases de données classique (type SQL) atteignent leur limite pour les interrogations
  - ✓ En quasi temps réel
  - ✓ Sur un grand volume de données et/ou une longue période de temps sur des données stockées à long-terme

afnic

Pourquoi une approche Big-Data ?

- ✓ Ambition d'utiliser jusque 100% du trafic DNS
- ✓ Mais…
  - ✓ Comment stocker autant de To de données brut
  - ✓ Comment traiter ces milliards de requêtes
  - ✓ Comment stocker les résultats des traitements
- ✓ Le Big-Data peut apporter des solution
- ✓ Mais…
  - ✓ Concrètement, lesquelles ?
  - ✓ Lesquelles seraient les plus pertinentes pour nous ?
  - ✓ Pas de compétences en interne au lancement de l'idée

afnic

Réalisation d'un PoC Big-Data (1/2)

- ✓ Le Big-Data était/est un buzzword
- ✓ Concepts introduits en 2001
- ✓ De nouvelles unités de mesure de stockage à connaître
  - ✓ Tera, Peta, Exa, Zetta,Yotta… octets
- ✓ Technologies en évolution constante.
  - ✓ 2012: La priorité est au stockage des données
  - ✓ 2014: La priorité est à l'accès temps-réel à ces données.
- ✓ Nécessité de créer une synergie aux nombreuses compétences.
  - ✓ Sysadmins, développeurs, data-scientist/data-analyst, ingénieurs métier, …

afnic

Réalisation d'un PoC Big-Data (2/2)

- ✓ Partenariat avec Citizen-Data
  - ✓ Expertise en solutions Big-Data
  - ✓ Proposent des solutions d'analyse de capture DNS (pcap)
- ✓ Choix d'une solution et premiers tests en interne
  - ✓ Framework Hadoop et distribution Hortonworks validés
    - ✓ Solution Logiciel libre (créé en 2006, écrit en Java)
    - ✓ Des poids lourds de l'Internet à l'origine (Google, Yahoo!, Facebook, Ebay…)
    - ✓ Utilisé par des homologues registres
    - ✓ Choix entériné par notre partenaire
- ✓ Location d'un « Cluster » le temps de réaliser le PoC
  - ✓ 5 serveurs pour les données et les traitements + 1 serveur C&C

afnic

Défis à relever/Attentes

- ✓ Défis
  - ✓ Adapter/re-développer nos outils de traitement (nouveaux langages/ concepts…)
  - ✓ Transporter les données des sondes vers le cluster
  - ✓ Proposer des traitements proche du temps-réel
    - ✓ Ne pas réaliser uniquement des analyses « Post-Mortem »
  - ✓ S'adapter à un framework en pleine phase de développement
- ✓ Attentes
  - ✓ Capacités de traitement proche de 100% du trafic DNS pour améliorer la pertinence des indicateurs de résilience
  - ✓ Transfert de connaissance pour intégration d'autres types de données internes, « fusion » avec la B.I.
  - ✓ Améliorer nos outils de lutte contre des attaques au DNS (ex: DDoS)

afnic

Perspectives et pistes

- ✓ De nouveaux indicateurs pour les prochaines éditions
  - ✓ Qualité technique des zones
  - ✓ Résolveurs les plus demandeurs
  - ✓ Et d'autres viendront compléter…
- ✓ De nouveaux outils
  - ✓ La plateforme Big-Data qui sera opérationnelle pour les mesures
  - ✓ Zonemaster (http://zonemaster.fr)
    - ✓ Outil développé par l'AFNIC et le .SE (http://zonemaster.fr)
    - ✓ Doit remplacer notre ancien outil « ZoneCheck »
    - ✓ https://github.com/dotse/zonemaster/wiki/Zonemaster-Distribution-Releases
- ✓ … et des sources complémentaires d'informations à exploiter
  - ✓ Réseau de sondes Atlas

afnic

Merci !

www.afnic.fr
contact@afnic.fr
Twitter : @AFNIC
Facebook : afnic.fr