RéSIST

Le filtrage de la navigation Internet : retour d'expérience

RéSIST

Contexte

- Enseignement supérieur
- 20000 personnes

RéSIST

Pourquoi le filtrage ?

- Pour des raisons de débit
  - 30% du débit par des sites non souhaitables
- Pour des raisons légales
  - Contenus illégaux
  - Présence de mineurs
  - Obligations pour les FAI
- Pour des raisons de productivité
- Pour des raisons de sécurité

RéSIST

Que filtrer ?

- Protection enfance
  - Pornographie, explosifs, drogue
- Législation
  - Révisionnisme, pédophilie, zoophilie
- Productivité
  - Presses, jeux, casino
- Sécurité
  - Spyware / antivirus.

RéSIST

Comment filtrer ?

- Tout est interdit sauf (liste blanche)
- Tout est autorisé sauf (liste noire)

RéSIST

Comment filtrer ?

- Notation ICRA
- Liste d ’URLs
- Analyse du contenu textuel
- Analyse d ’images (pornographie)

RéSIST

Notation ICRA

- Les gentils fournisseurs de sites de pornographie vont dire qu ’ils fournissent de la pornographie pour pouvoir être bloqués. (1 pour 300)

RéSIST

Liste d ’URLs

- Principe
  - base d ’urls et/ou d ’IP à bloquer
  - trésor de guerre des sociétés de filtrage
- Avantages
  - Très rapide (15000 req/s)
  - Facile à corriger
  - Très simple
- Inconvénients
  - ne filtre que ce qu ’elle connaît (principe des signatures)

RéSIST

Liste d ’URLs

- Les questions à se poser :
  - Combien de catégories ? (plus = mieux ?)
  - Régularité de mise à jour
  - Contexte de génération
- Exemple
  - squidguard
  - privosquid (Windows)
  - La quasi totalité des logiciels commerciaux

RéSIST

Analyse du contenu textuel

- Principe :
  - mots clés / expressions régulières dans la page
  - Souvent en complément du filtrage d ’urls
- Avantages
  - Détecte même l ’inconnu
  - Filtrage plus complet

RéSIST

Analyse du contenu textuel

- Inconvénients
  - Lent
- La qualité de la détection des schémas textuels est fondamental
  - Qualité et rapidité souvent antinomique

RéSIST

Analyse du contenu textuel

- Les questions à se poser
  - Qualité du détecteur
- Exemples :
  - dansguardian (1 seconde maximum par page)
  - DWP de dolphian
  - privoxy (ou privosquid)

RéSIST

Analyse des images

- Principe
  - Pages de type érotique / pornographique
  - Proportion colorimétrique
  - Reconnaissance de formes (?!?!)
- Avantages
  - Page sans texte.
- Inconvénients
  - Lenteur (10s pour certaines pages !)
- Exemples
  - Pure-Sight

RéSIST

Les questions à se poser

- Quel est mon contexte ?
  - Université/recherche = large ouverture
  - Entreprise = ouverture moyenne à faible
  - Ecole/maison = ouverture faible
- Quels sacrifices j ’accepte de faire ?
  - Quel taux de faux positifs ?
  - Quel taux de faux négatifs ?
  - Quel ralentissement ?

RéSIST

Les questions à se poser (2) ?

- Où est-ce que je place mon filtrage ?
  - Sur le poste ? Plus efficace car plus de CPU disponible
  - En coupure sur le réseau ? non contournable, centralisé
- Fonctionne avec ICAP (échange de contenu) ?
- Procédures de contournement ?
- Filtrage du HTTPS

RéSIST

Les questions à se poser (3) ?

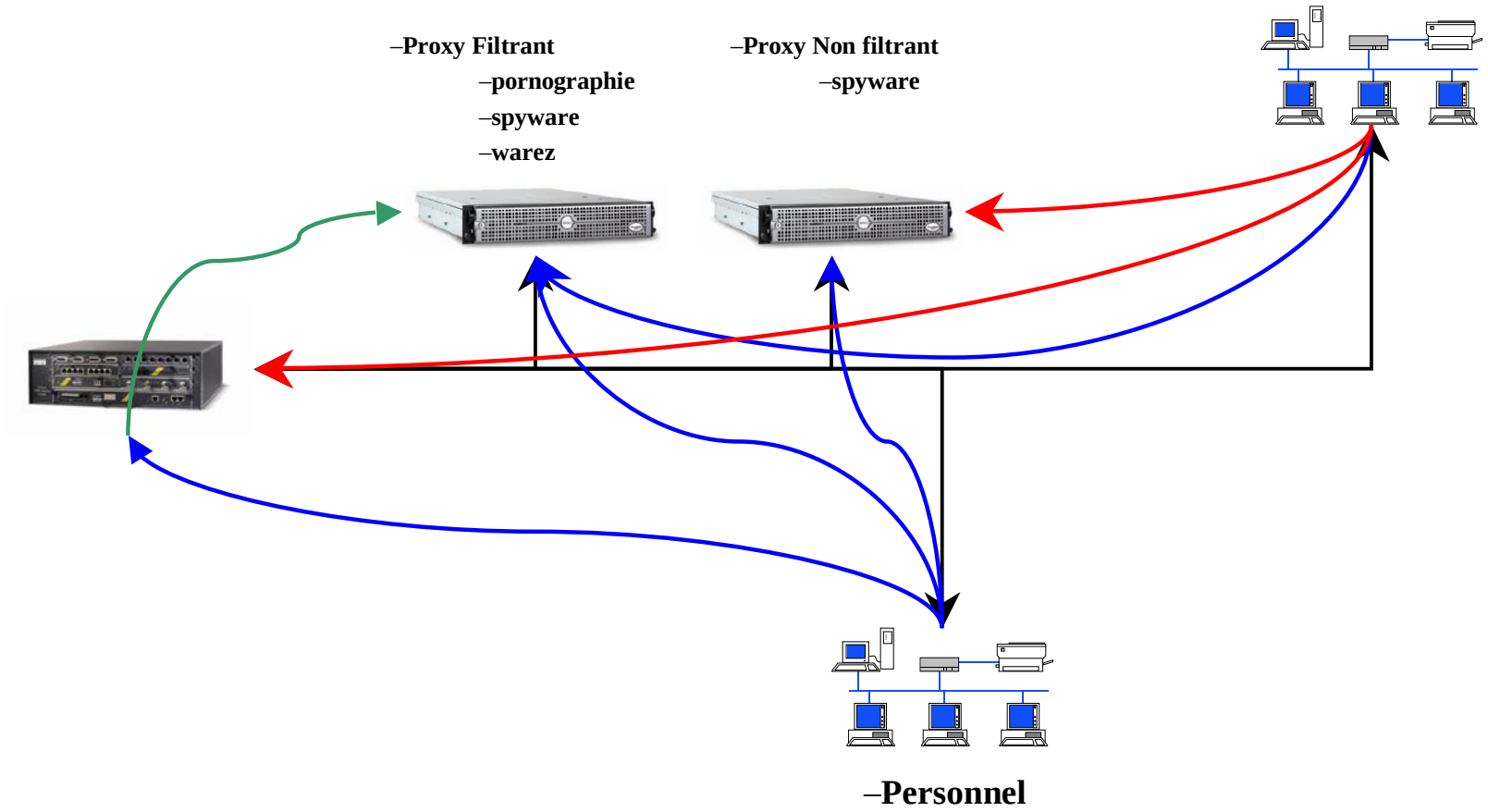
- Responsabilisation du « contrevenant »?
- Profil des utilisateurs (par IP, par user, etc..)
- Filtrage Horaire
- Administrable en direct par script ?

RéSIST

A l ’UT1

- 2 proxy : 1 filtrant + 1 non filtrant
- Squid + squidguard + bases internes (604000)
- Utilisation de bases « spyware »
- 12 Listes noires
- 1 Liste blanche

A l ’UT1


–Etudiants
–Proxy Filtrant
–pornographie
–spyware
–warez
–Proxy Non filtrant
–spyware
–Personnel

RéSIST

Politique

- 1 proxy filtrant pour tout le monde
- 1 proxy non filtrant pour les personnels (pas par défaut)
- 1 proxy transparent pour les petits malins
- Sont automatiquement bloqués
  - le pornographique
  - les spywares (base surbl.org)
  - les warez sont bloqués

RéSIST

Politique (2)

- Le filtrage est plus strict pour les bibliothèques et certaines zones
  - liste blanche
  - restrictions sur les webmail, etc.
- Certaines salles sont complètement verrouillées l ’après-midi

RéSIST

Résultats pour 1 semaine

- Blocages
  - 8876 blocages adultes
  - 3960 redirecteurs (majoritairement google-images)
  - 81 blocages de spyware
  - 13 warez
  - 2 phishing

RéSIST

Résultats pour 1 semaine

- 10 demandes d ’ajout à la liste blanche
- 1 plainte par semaine
- 50-200 ajouts dans la base par jour

RéSIST

Filtrage web : détection

- Les filtres détectent les abus pornographiques
  - volontaires (on s ’en moque)
  - involontaires => spyware modifiant les pages d ’accueil => Virus
- Les filtres détectent les sorties spyware
  - Poste infecté
- Les filtres détectent les sorties « updates virales »
  - Poste infecté

Infection

Marketingware

Sites pornographiques

Spyware

Récupération des données

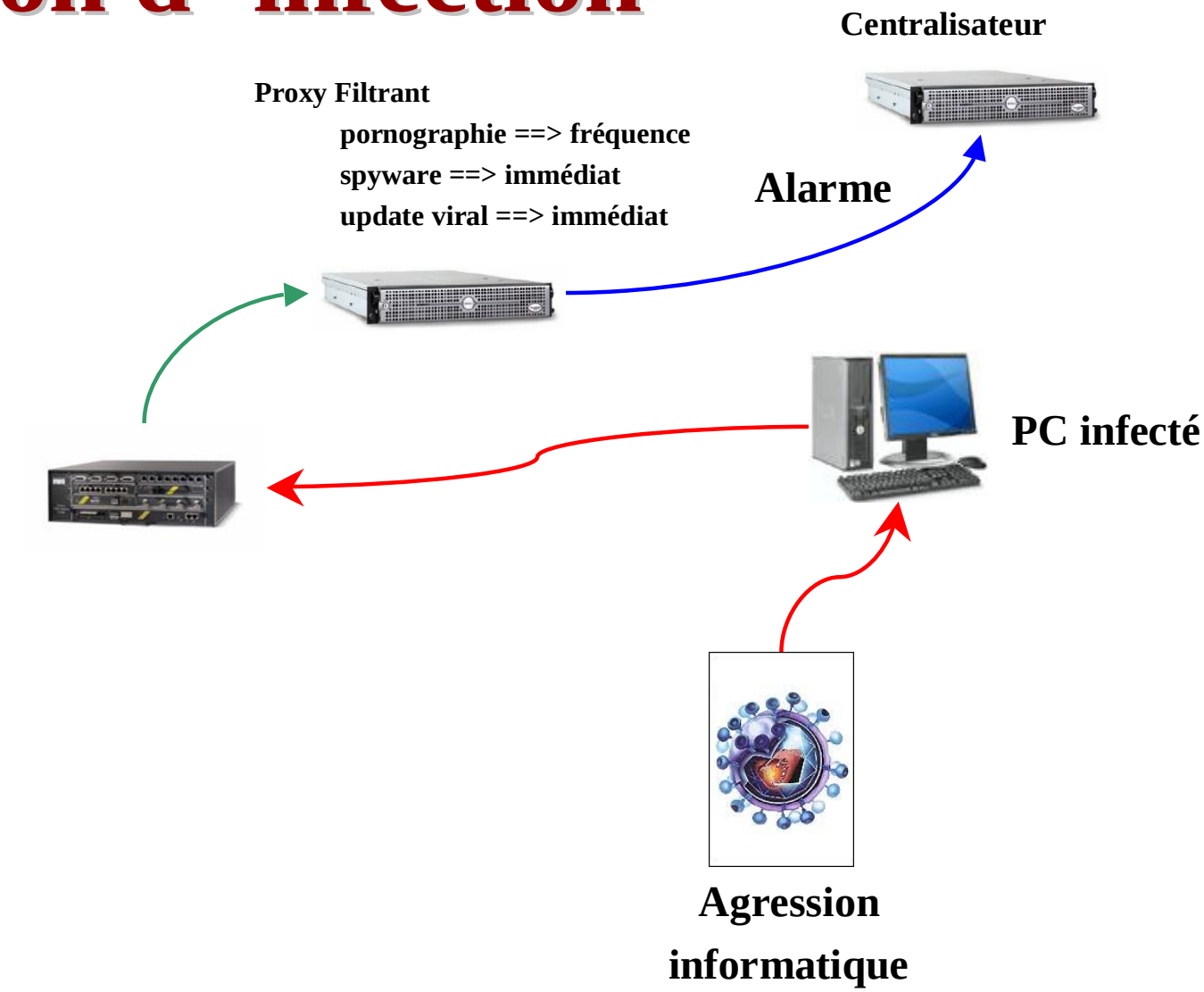
Virus / Ver

Mise à jour Virales

PC infecté

Agression informatique

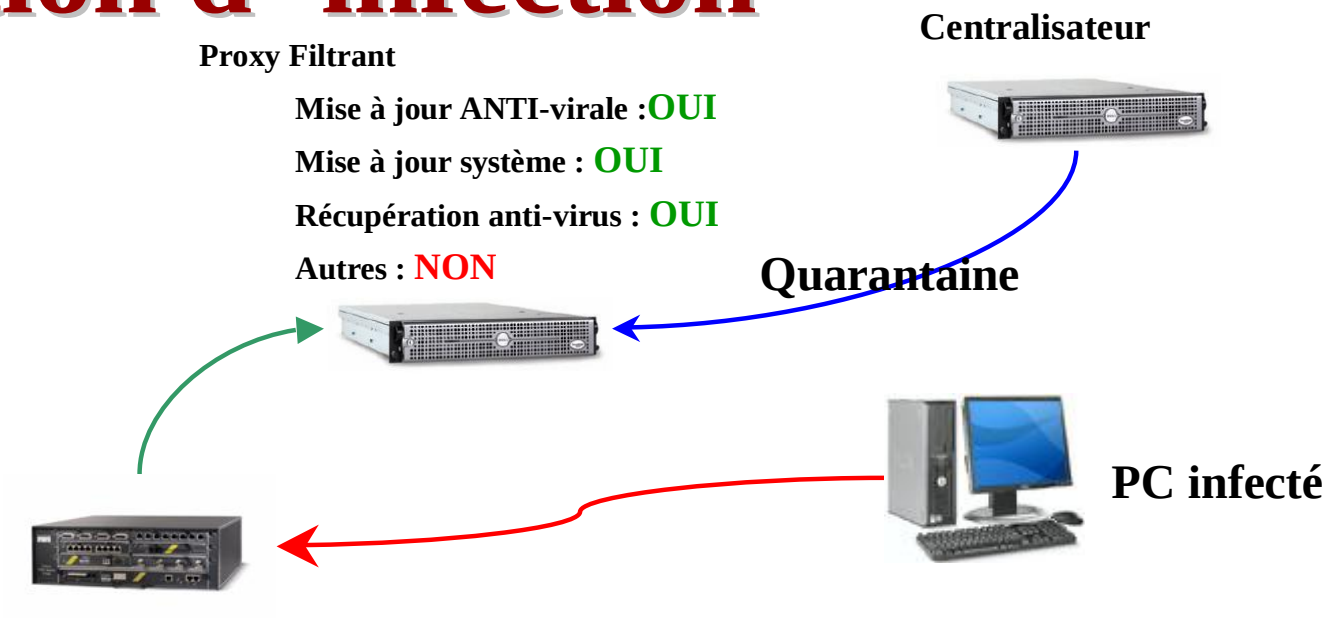
Détection d ’infection



Filtrage web : action

- Les postes infectés sont dans une liste
- Le filtrage web bloque les IP de cette liste hormis base cleaning
  - Impossibilité de mise à jour du virus
  - Avertissement clair et immédiat de l ’utilisateur
  - Une procédure de désinfection est proposée avec la possibilité de mettre à jour antivirus et logiciels

Limitation d ’infection



RéSIST

Résultats pour 1 semaine

- 81 blocages de spyware
- 2 phishing
- 20 postes infectés (dont 15 WiFi) bloqués

RéSIST

Conclusions

- Le filtrage « administratif » est efficace
  - Les salles sont correctement utilisées (avec félicitations de certains utilisateurs)
  - Les plaintes sont très faibles
- Le filtrage sécuritaire est efficace
  - Le diagnostic est immédiat et clair
  - Les utilisateurs de PC infectés sont agréables (ils sont fautifs, et ils le savent)

RéSIST

Conclusions (2)

- Filtrage URL adapté à **notre** environnement
- Résultats forcément différents
  - Pour une école.
  - Pour un particulier
  - Pour une entreprise