CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Le monitoring de flux réseaux à l'IN2P3 avec EXTRA

Journée JoSy « Supervision systèmes et réseaux dans un laboratoire de recherche »

27 mars 2008, ENS Paris

Denis Pugnère, CNRS / IN2P3 / IPNL

basé sur une présentation de J.Fulachier aux JI/IN2P3-CEA - Hourtin - 2004

Plan de la présentation

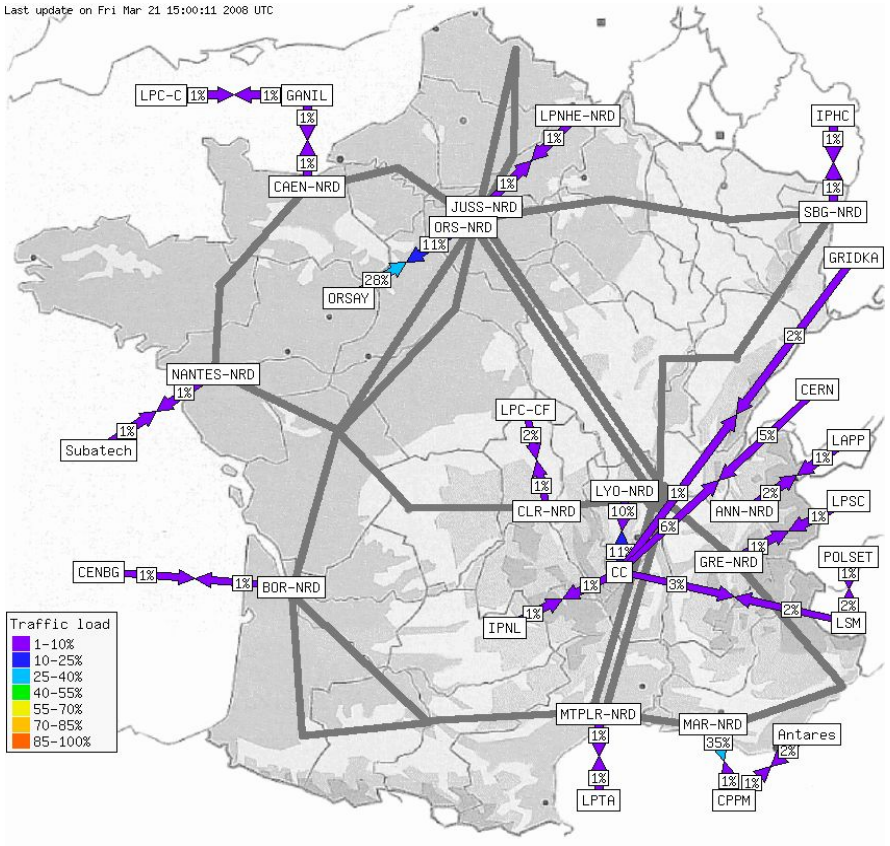
1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
5. Contributeurs et liens

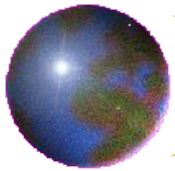
L'IN2P3 en chiffres

- Institut National de Physique Nucléaire et de Physique des Particules
- 2565 personnes (chercheurs et ITA)
- 20 laboratoires et structures : APC, CENBG, Centre de Calcul, CPPM, CSNSM, IMNC, IPHC, IPNL, IPNO, Ganil, LAL, LAPP, LLR, LMA, LPC Caen, LPC Clermont, LPNHE, LPSC, LPTA, LSM, Subatech, ULISSE

Le centre de calcul de l'IN2P3 en bref

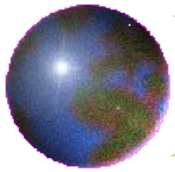
- Le CC-IN2P3 :
  - assure la connectivité de l'ensemble des laboratoires,
  - héberge le NRD RENATER à LYON,
    - héberge des moyens de calcul importants (> 2300 processeurs),
    - héberge des moyens de stockage importants,
    - Tier-1 LCG





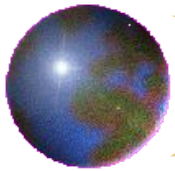
Plan de la présentation

1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
5. Contributeurs et liens



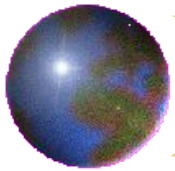
Cahier des charges de la métrologie réseaux à l'IN2P3

- Installer une solution de métrologie à tous les laboratoire de l'IN2P3 :
  - Pour satisfaire aux exigences de la loi :
    - Traçabilité des connexions,
    - Durée de conservation des traces,
  - Détection de problèmes de sécurité (volumes anormaux, détection de scans),
  - Permettre aux CSSI des laboratoire d'accéder à la métrologie réseau de leur laboratoire,
  - Permettre au Chargé de Mission SSI de l'IN2P3 d'accéder à la métrologie réseau de l'ensemble des laboratoires,
  - Analyse fine du trafic,
  - Administration simplifiée et centralisée de l'application,
  - Archivage centralisé des flux pendant la durée légale



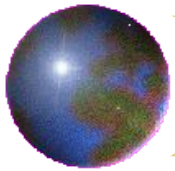
Plan de la présentation

1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
5. Contributeurs et liens



EXTRA ?

- Extra : EXternal TRafic Analysis
- Développé au LPSC
- Logiciel d'analyse des flux réseaux :
  - Basé sur : une sonde netflow, serveur TOMCAT et le code d'Extra en Java, le framework AMI, une base MySQL,
  - Récupération et stockage des flux réseaux dans une base de données,
  - Pré-traitements systématiques sur les flux stockés,
  - Interface graphique d'analyse (certificats CNRS-Standard),
  - Analyse fine : Analyse par adresse, protocole, port, volume, détection de propagation de vers, détection de volumes anormaux



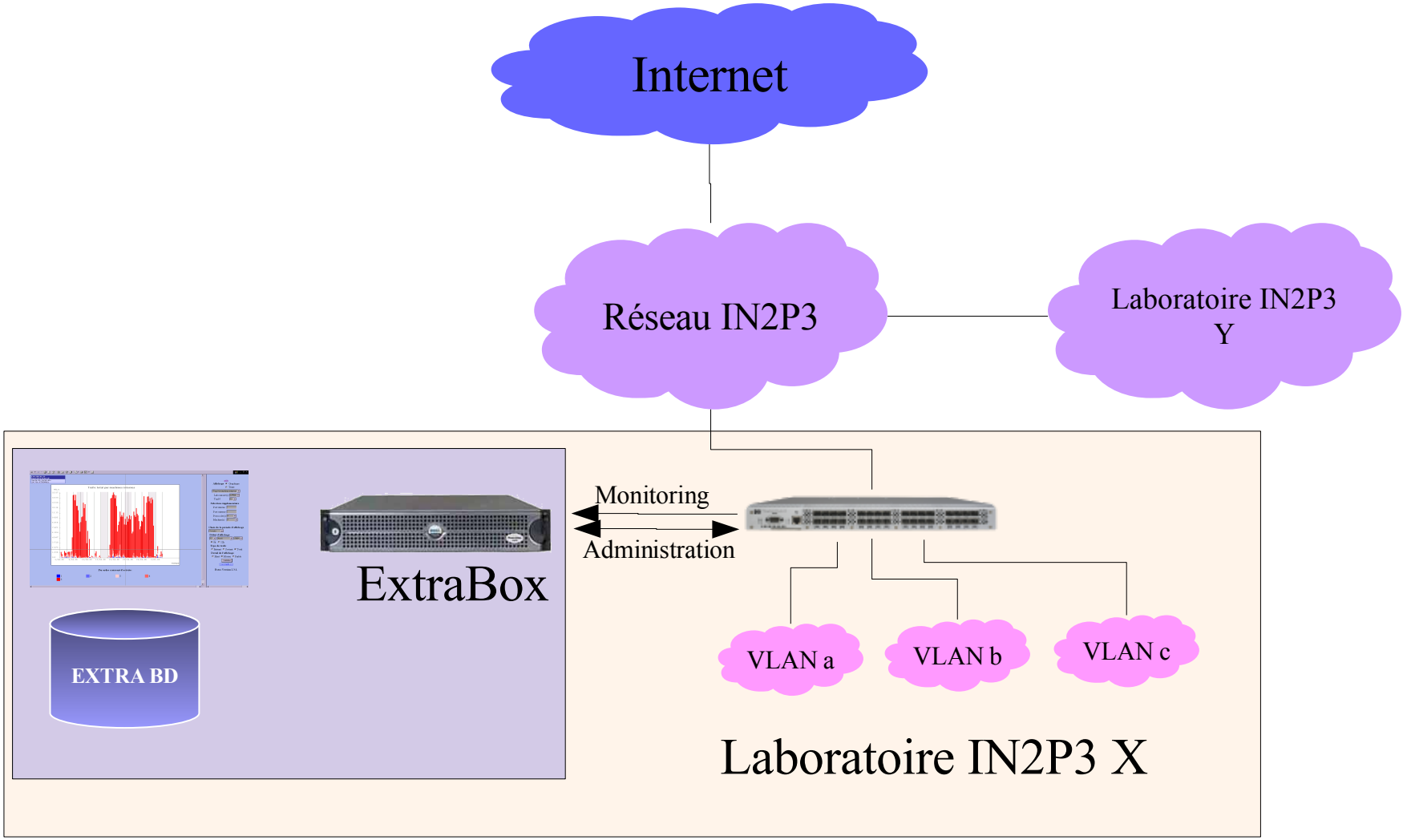
Déploiement d'EXTRA à l'IN2P3

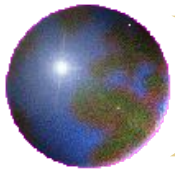
- Une « ExtraBox » fournie par le CC à chaque laboratoire (Serveur 1U, bi-xeon 3,4Ghz, 2 interfaces réseau Gigabit, 1Go RAM, disque 160Go)
- Fourniture d'un Système auto-installable sur CD « Bundle-Extra » :
  - À base de Scientific Linux, contient tous les utilitaires et librairies nécessaires à l'application EXTRA,
  - Détection disque dur, auto-partitionnement disque,
  - Configuration réseau in-situ (dans chaque laboratoire),
- Déploiement du logiciel EXTRA et de la configuration depuis le CC,
- Récupération des traces journalières

Installation de L'ExtraBox

- Phase 1 : Installation
  - Installation du serveur par l'ASR du laboratoire (port ethernet d'administration, port ethernet de monitoring),
  - Démarrage sur le CD « Bundle-Extra » : configuration réseau, auto installation,
- Phase 2 : Déploiement
  - Déploiement et configuration du logiciel EXTRA, installation des certificats (des CSSI et du serveur), activation du logiciel
- Phase 3 : Production
  - Récupération périodique des traces,
  - Mises à jour d'EXTRA (déployées depuis le CC) ou du Bundle-Extra (envoi d'un nouveau CD)

L'ExtraBox dans un laboratoire IN2P3

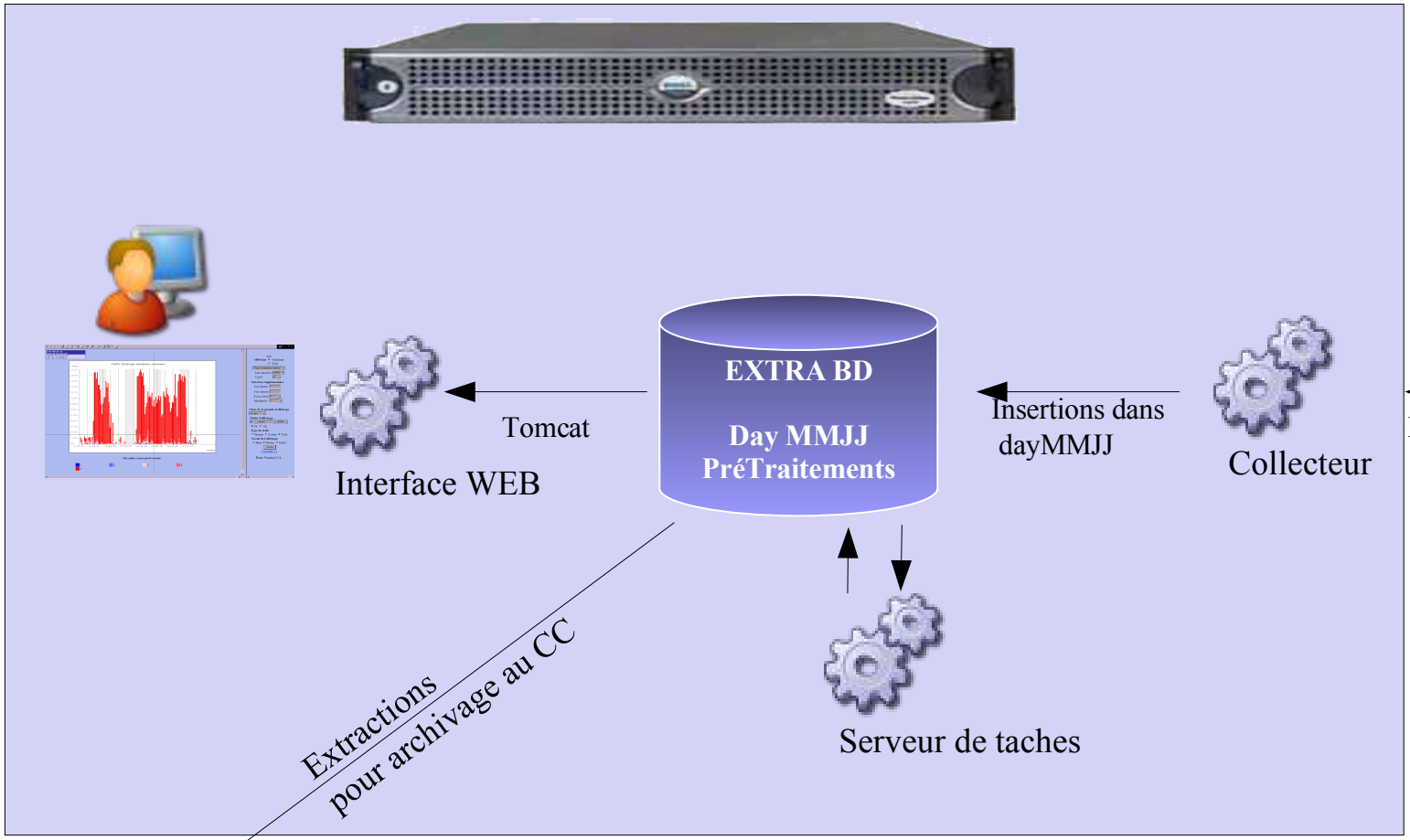


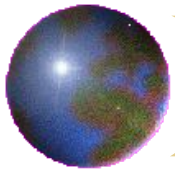


Plan de la présentation

1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
    1. Le collecteur
    2. Le serveur de tâches
    3. L'interface Graphique
    4. Les requêtes centralisées
5. Contributeurs et liens

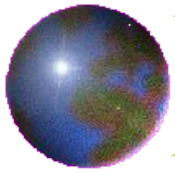
Architecture d'EXTRA





Plan de la présentation

1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
    1. Le collecteur
    2. Le serveur de tâches
    3. L'interface graphique
    4. Les requêtes centralisées
5. Contributeurs et liens

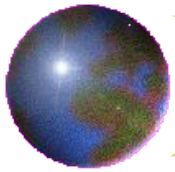


Le Collecteur d'EXTRA

- Il récupère les flux en provenance :
  - du routeur à condition qu'il supporte les netflows,
  - d'une sonde nprobe
- Il permet de moduler l'agrégation des flux.
- Il les met en forme.

  IPSource - IPDest - portSource - portDest - volume

  IPinterne - IPexterne - portInterne - portExterne - sens - volume
- Il insère les flux dans la base de donnée.

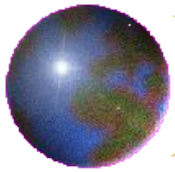


Plan de la présentation

1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
    1. Le collecteur
    2. Le serveur de tâches
    3. L'interface graphique
    4. Les requêtes centralisées
5. Contributeurs et liens

Le Serveur de tâches

- Lance des pré-traitements au fur et à mesure de l'arrivée des flux dans la base de données : accélérer la visualisation lors de l'analyse
- Permet d'accéder instantanément aux TopN, exemples :
  - Top ten des machines internes ou externe ayant généré le trafic le plus important.
  - Top ten des services internes les plus utilisés depuis l'extérieur du site.
  - Top ten des services externes les plus utilisés depuis l'intérieur du site.
- Permet l'ajout de nouveaux pré-traitements générés par les CSSI, exemple : surveillance du trafic d'un serveur web, mail, ssh...

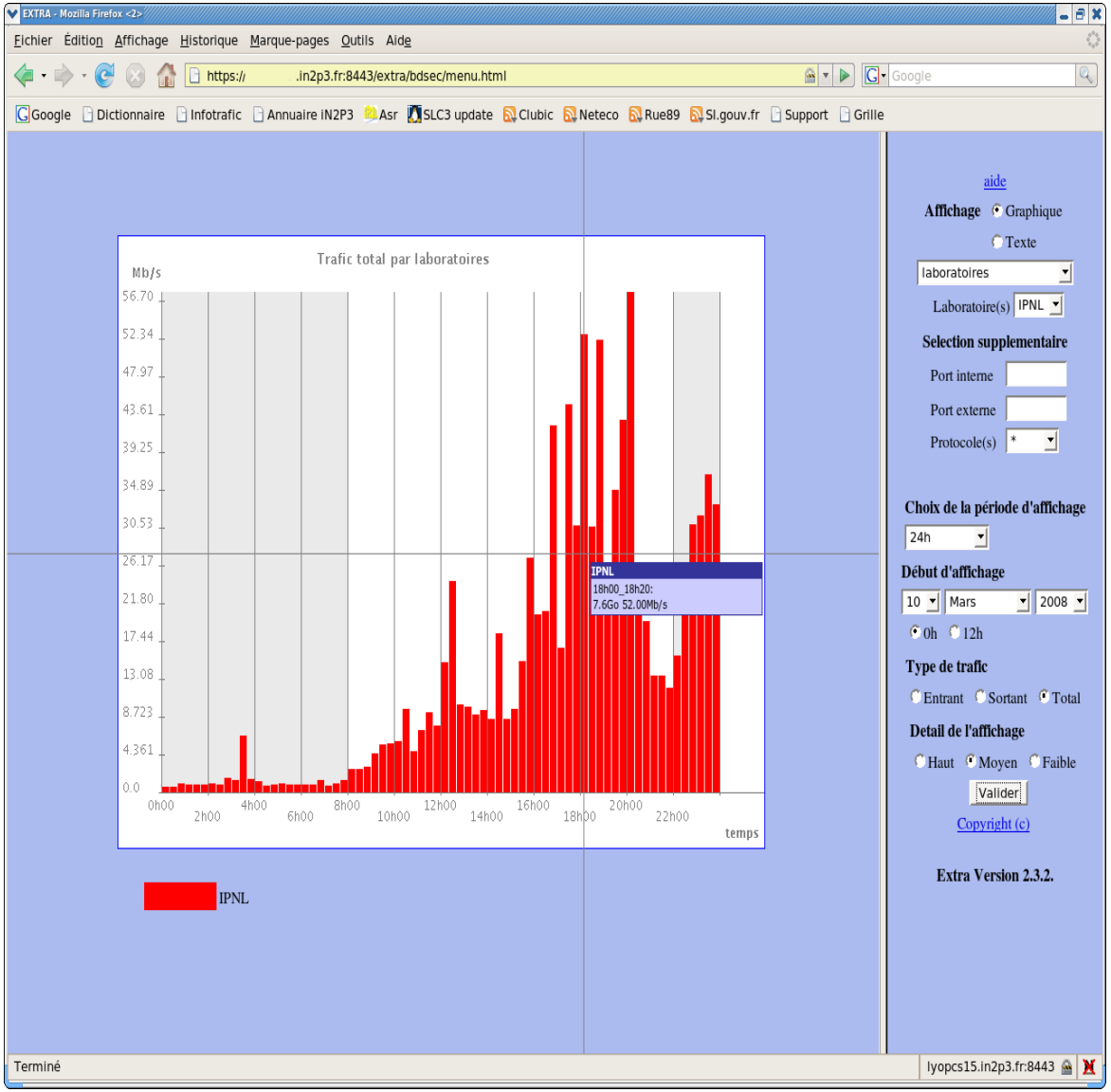


Plan de la présentation

1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
    1. Le collecteur
    2. Le serveur de tâches
    3. L'interface graphique
    4. Les requêtes centralisées
5. Contributeurs et liens

Fonctionnalités générales de l'interface WEB

- Visualisation du trafic :
  - général, par port, par machine,
  - Top 1 à 10
- Sélection du trafic : par machine, par subnet, part port (interne ou externe), par protocole,
- Sélection du temps et de la période,
- Sélection du type de trafic : entrant, sortant, total
- Détail de l'affichage,
- Possibilité de saisir des requêtes manuelles en SQL,
- Manuel en ligne intégré à l'application



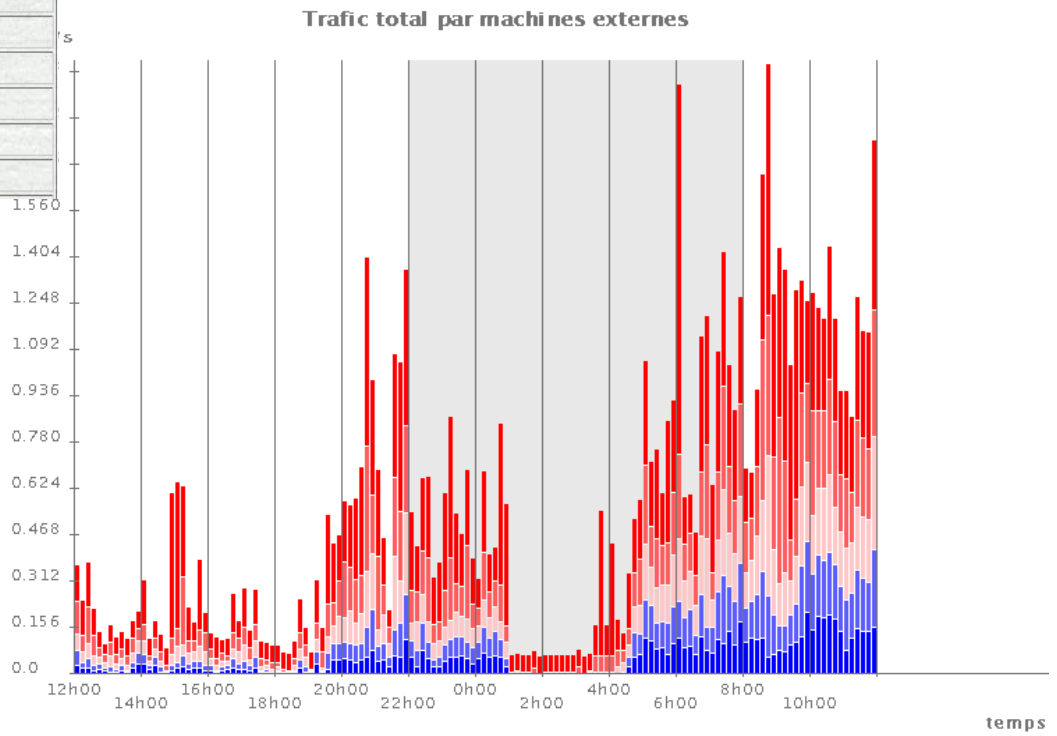
Interface WEB : visualisation

Trafic total par machines externes

periode	machine	volume
12h00_12h10	207.228.239.4	405.8Ko
12h00_12h10	134.158.69.191	424.7Ko
12h00_12h10	217.83.126.166	1.6Mo
12h00_12h10	62.211.194.218	2.1Mo
12h00_12h10	217.120.63.121	4.0Mo
12h00_12h10	213.115.152.12	4.3Mo
12h00_12h10	205.251.183.137	8.4Mo
12h00_12h10	12.231.53.164	8.9Mo
12h10_12h20	193.253.62.4	509.1Ko
12h10_12h20	193.51.65.71	573.0Ko
12h10_12h20	217.83.126.166	1.4Mo
12h10_12h20	62.211.194.218	1.4Mo

Sous forme texte

Sous forme graphique



Interface WEB : sélection du trafic

- Les différentes vues possibles :
  - Tri par machines internes : port interne, machine externe, port externe, volume
  - Tri par machines externes : port externe, machine interne, port externe, volume
  - Tri par ports interne : machine interne,machine externe, port externe, volume
  - Tri par ports externes : machine externe,machine interne, port interne, volume
  - Services prédéfinis : service, port, protocole, machine, volume
  - Notion de top N pour les tris.

Interface WEB : sélection des paramètres

- En fonction du TopN choisi : Sélection optionnelle de paramètres supplémentaires
  - Port interne ou externe
  - Protocole : TCP, UDP, ICMP, tous
  - Machine : toutes, subnet, une seule (nom ou adresse IP)

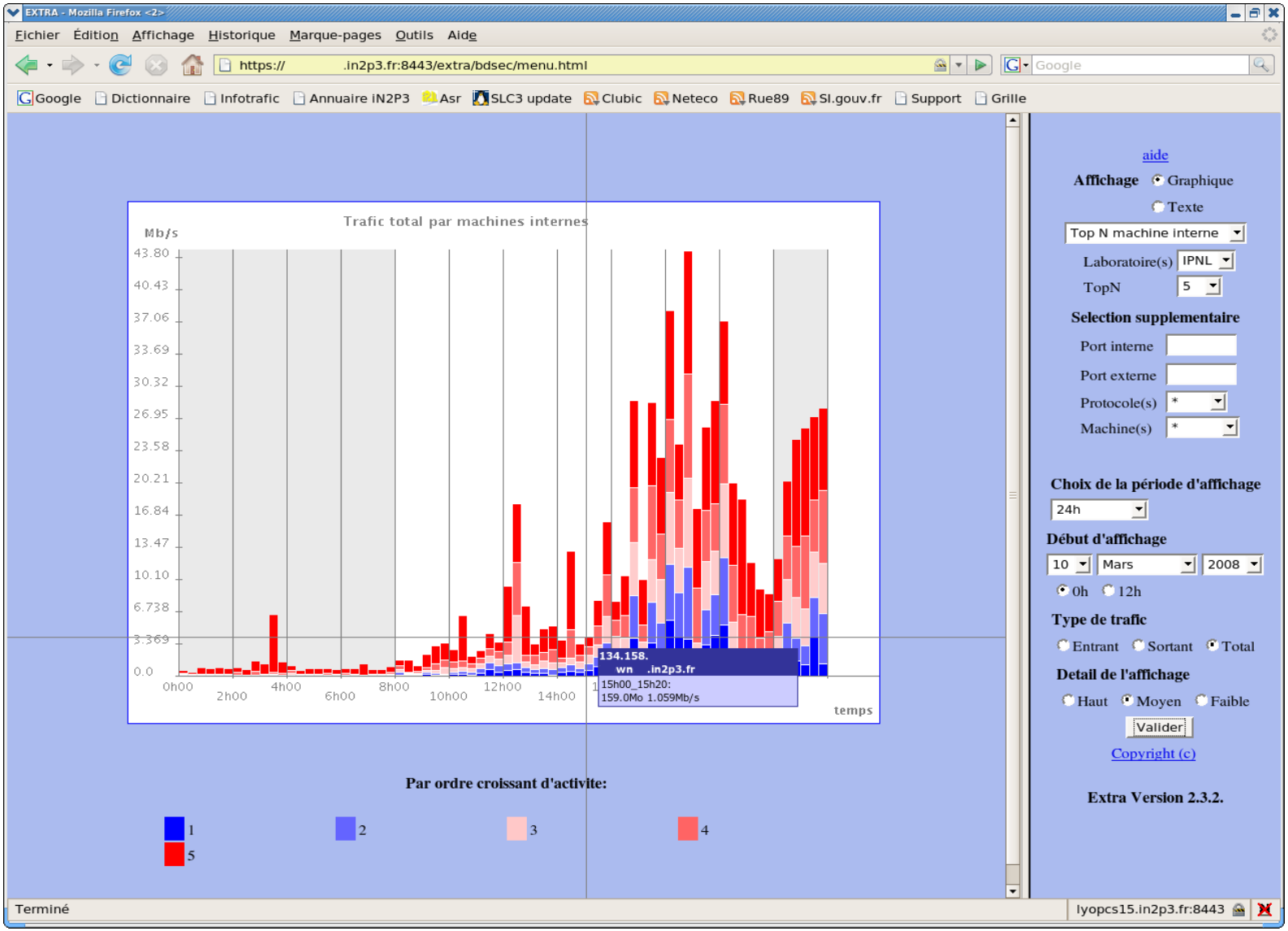
Interface WEB : choix d'une période d'affichage

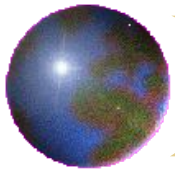
- Période :
  - 24h,
  - 48h,
  - 1 semaine,
  - 1 mois
- Début d'affichage : au choix dans les 3 mois d'historique en ligne

Interface WEB : choix du type de trafic

- Sélection du trafic
  - Entrant dans le laboratoire
  - Sortant du laboratoire
  - Total

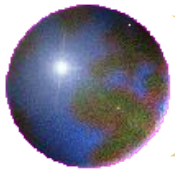
Interface WEB : résultat de la sélection





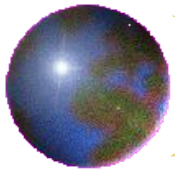
Plan de la présentation

1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
    1. Le collecteur
    2. Le serveur de tâches
    3. L'interface graphique
    4. Les requêtes centralisées
5. Contributeurs et liens



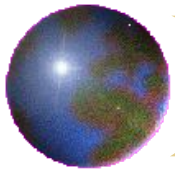
Requêtes centralisées

- Requêtes planifiées :
  - Détection de scan d'une machine interne (exemple : infectée par un virus),
  - d'archivage,
  - de maintenance
- Requêtes spécifiques
  - Recherche de connexions d'une adresse IP spécifique sur tous les laboratoires (exemple : suite à une alerte d'un CERT ou d'un site distant)



Plan de la présentation

1. Introduction
2. Cahier des charges de la métrologie réseau IN2P3
3. Le logiciel EXTRA et son déploiement
4. Fonctionnalités générales
    1. Le collecteur
    2. Le serveur de tâches
    3. L'interface Web
    4. Les requêtes centralisées
5. Contributeurs et liens



Contributeurs et liens

- Contributeurs au projet : Bernard Boutherin (LPSC/IN2P3), Laurent Caillat-Vallet (CC-IN2P3), Jérôme Fulachier (LPSC/IN2P3), Jean Mirasolo (LPSC/IN2P3), Denis Pugnère (IPNL/IN2P3)
- Liens :
  - Poster JRES :
    Résumé : http://2005.jres.org/resume/poster/6.pdf
    Poster : http://2005.jres.org/poster/6-boutherin.pdf
  - Logiciel EXTRA :
    http://lpsc.in2p3.fr/extra/
  - Framework AMI (ATLAS Metadata Interface) :
    https://twiki.cern.ch/twiki/bin/view/Atlas/AtlasMetadataInterface
  - RAMUX (par Serge Bordères) :
    http://www.cenbg.in2p3.fr/article.php3?id_article=245