

Logiciel libre, une introduction

Roberto Di Cosmo



Université Paris Diderot
UFR Informatique
Laboratoire Preuves, Programmes et Systèmes
roberto@dicosmo.org

April 3, 2012

Génie Logiciel

Linux
Eric S. Raymond
Apache
Mythes et réalité

Distributions Linux

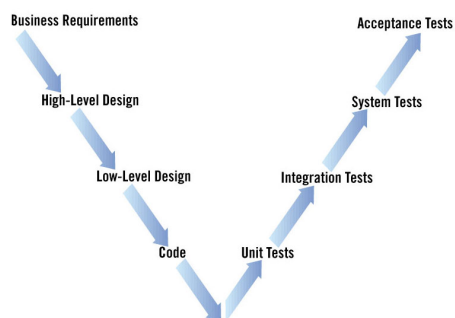
Theoretical results

Securité

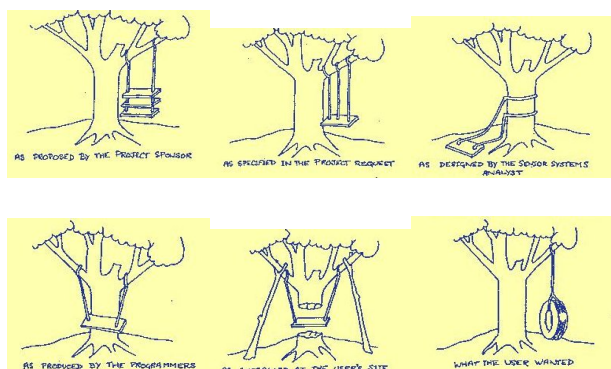
Exemples du monde propriétaire
Vote électronique

Un peu d'histoire du Génie logiciel

Le "modèle à V" incarne la vision du développement logiciel typique des années '80



Le problème en six images



Part V

Génie Logiciel Libre

Génie Logiciel et Logiciel Libre

Des origines à nos jours

Défauts du modèle

- ▶ il faut une organisation centralisée
- ▶ les spécifications initiales peuvent être fausses ou incomplètes
- ▶ la distance entre spécifications et test final peut être trop longue
- ▶ l'utilisateur final est exclu du processus jusqu'à la fin
- ▶ quand on découvre un erreur grave, il est souvent trop tard

Des solutions proposées ensuite...

- ▶ programmation orientée objet
- ▶ conception et modélisation via UML
- ▶ utilisation de méthodes formelles dans le développement
- ▶ diverses méthodologies plus récentes:
 - ▶ extreme programming
 - ▶ méthodes agiles
 - ▶ ...

Voyons voir concrètement ce qui s'est passé avec certains projets en logiciel libre.

Analyses

Exemples de communautés significatives

- ▶ Linux
- ▶ Apache

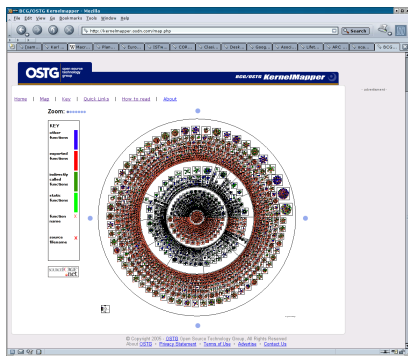
Exemple de patch dans la mailing list

```
Date: Fri, 16 Mar 2007 12:36:15 +0100
From: Jarek Poplawski <>
Subject: [PATCH] Re: Fwd: oprofile lockdep warning on rcl

On 28-02-2007 01:46, Dave Jones wrote:
> This happened on a 2.6.21rc1 kernel.
...
> Richard Hughes <hughesient@gmail.com>
> Date:
> Tue, 27 Feb 2007 21:59:07 +0000
> To:
> Development discussions related to Fedora Core
> <fedora-devel-list@redhat.com>
...
> But also I get this (!!!):
>
Here is my patch proposal for testing.

Regards,
Jarek P.
```

Un aperçu: structuration



Analyses

Un texte très connu tire des leçons du succès de Linux :

The cathedral and the Bazaar, Eric S. Raymond, 1997

- cathedral : processus logiciel lourd et hiérarchique
- bazaar : développement par des équipes avec faible couplage

Linux: une radiographie

- ▶ quelques dates:
vers. 0.02 en 10/1991; 0.95 en 03/1992; 1.0 en 03/1994
- ▶ modèle du *dictateur bienveillant*:
Linus Torvalds a le dernier mot sur tout
- ▶ élaboration des patches sur la mailing list (cela permet de discuter les propositions)
- ▶ séparation entre fils de versions:
stable (2.4.x, 2.6.x), et *experimental* (1.3.xx or 2.1.x)
- ▶ gestion de version avancée (plus de détails dans un cours à part)

Exemple de patch dans la mailing list

```
lockdep found oprofilefs_lock is taken both in process
context (oprofilefs_ULONG_from_user()) and from hardirq
(mmi_cpu_setup()), so the lockup is possible.

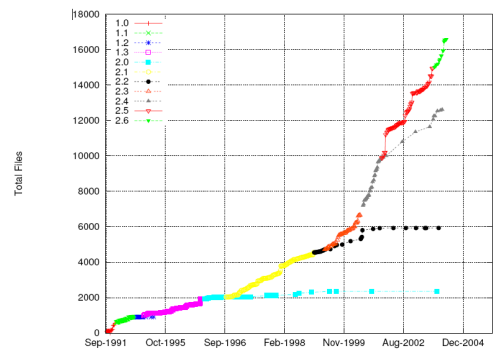
Reported-by: Richard Hughes <hughesient@gmail.com>
Signed-off-by: Jarek Poplawski <jarkao2@o2.pl>

---
diff -Nurp linux-2.6.21-rc1-/drivers/oprofile/oprofilefs.c linux-2.6.21-rc1-/drivers/oprofile/oprofilefs.c
--- linux-2.6.21-rc1-/drivers/oprofile/oprofilefs.c 2007-02-27 10:47:38.000000000 +0100
+++ linux-2.6.21-rc1-/drivers/oprofile/oprofilefs.c 2007-03-16 11:36:30.000000000 +0100
@@ -65,6 +65,7 @@ ssize_t oprofilefs_ULONG_to_user(unsigned
int oprofilefs_ULONG_from_user(unsigned long * val, char const __user * buf, size_t count)
{
    char tmpbuf[TMPBUFSIZE];
+   unsigned long flags;

    if (!count)
        return 0;
@@ -77,9 +78,9 @@ int oprofilefs_ULONG_from_user(unsigned
if (copy_from_user(tmpbuf, buf, count))
    return -EFAULT;

-   spin_lock(&oprofilefs_lock);
+   spin_lock_irqsave(&oprofilefs_lock, flags);
+   *val = simple_strtol(tmpbuf, NULL, 0);
-   spin_unlock(&oprofilefs_lock);
+   spin_unlock_irqrestore(&oprofilefs_lock, flags);
    return 0;
}
```

Un aperçu: évolution



Évolution du nombre des fichiers dans le noyau Linux (G. Robles)

Règles de bonne conduite énoncées

1. Every good work of software starts by scratching a developer's personal itch.
2. Good programmers know what to write. Great ones know what to rewrite (and reuse).
3. "Plan to throw one away; you will, anyhow." (Fred Brooks, The Mythical Man-Month, Chapter 11)
4. If you have the right attitude, interesting problems will find you.
5. When you lose interest in a program, your last duty to it is to hand it off to a competent successor.
6. Treat your users as co-developers.
7. Release early. Release often.
8. Given enough eyeballs, all bugs are shallow.

Critiques

- ▶ analyse très orientée Linux
- ▶ separation entre Bazaar et Cathedral moins nette dans la réalité
- ▶ beaucoup d'approximations, par exemple, la dénégation de la loi de Brooks, qui oublie la différence entre core developers et utilis-acteurs à la marge du projet

Apache

- ▶ quelques dates:
Création du projet en 1995 par Rob McCool, première version stable en 1996, version 1.3 en 1998, IBM choisit Apache pour WebSphere en juin 1998
- ▶ approx. 1500 committers sur les projets (une centaine sur httpd)
- ▶ organisation *oligarchique*

*Committers get a shot at working on the code; good committers become members and thus get a piece of the ownership of the software and the direction. Commit access is a privilege, not a right, and is based on trust. The Apache Software Foundation is a meritocracy [. . .]
New candidates for membership are nominated by an existing member and then put to vote; a majority of the existing membership must approve a candidate in order to the candidate to be accepted.*

Reverse Software Engineering

On peut essayer d'analyser ces cas *systématiquement*:

- ▶ disponibilité des sources
- ▶ accès aux données historiques
- ▶ informations complètes sur les développeurs
- ▶ pas besoin d'accords formels

Un travail d'envergure a été entrepris par <http://libresoft.urjc.es/index>

Exemples d'architectures

Parenthèse: Free Software Lessons/Lean Software principles

Value

Every good work of software starts by scratching a developer's personal itch.

Customer first

Treating your users as co-developers is your least-hassle route to rapid code improvement and effective debugging.

Removing waste

Good programmers know what to write. Great ones know what to rewrite (and reuse).

Perfection, well. . . fix it fast

Release early, release often.
Given enough eyeballs, all bugs are shallow.

Lessons from E. S. Raymond's "The Cathedral and the Bazaar", 1997

Quelques mythes sur le Logiciel Libre

- ▶ développement anarchique sans leader
- ▶ connaissance qui emerge naturellement de la sagesse des masses
The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations
2004 James Surowiecki
- ▶ le logiciel libre est *toujours* de meilleure qualité que le logiciel propriétaire
- ▶ l'écosystème du logiciel libre est basé sur des valeurs d'altruisme communautaire. . .

Un point de vue moderne

Les analyses de Martin Michlmayr (ex Debian project leader) sont disponibles sur <http://opensource.mit.edu>

Cathedral phase Transition phase Bazaar phase

Original "idea"
Project Author
Core developers
Unix philosophy

⇒ "Interest"
Prototype
Modular design

⇒ Distributed development environment
Community
Parallel perfective and corrective maintenance
Peer reviews

Focus sur la modularité

Un livre récent (2011) édité par Amy Brown et Greg Wilson, permet de se faire une idée de l'importance de la *modularité*.

The Architecture of Open Source Applications recueille une description de l'architecture de 25 projets Logiciel Libre, faite par quelques uns de leurs auteurs.

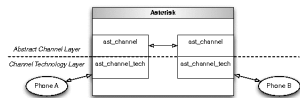
Voyons quelques exemples.

Asterisk (Russel Bryant)

Asterisk is...

an open source telephony applications platform distributed under the GPLv2.

Architecture basée sur l'abstraction des Channels et Channel Drivers:



Grâce à elle, un contributeur peut:

- ▶ fournir un nouveau pilote pour un nouveau type de téléphone sans se soucier du rester
- ▶ fournir une nouvelle fonctionnalité dans Asterisk sans se soucier des pilotes

Architecture extrêmement modulaire, qui a du succès depuis plus de 10 ans.

Eclipse (Kim Moir)

Eclipse is...

an open source platform for the creation of interoperable tools for application developers.

Eclipse's achievement

"Implementing software modularity is a notoriously difficult task. Interoperability with a large code base written by a diverse community is also difficult to manage. At Eclipse, we have managed to succeed on both counts."

Eclipse Architecture

Les composants essentiels sont:

three major elements, which corresponded to three major sub-projects: the Platform, the JDT (Java Development Tools) and the PDE (Plug-in Development Environment).

Observation

In order to encourage people to build upon the Eclipse platform, there needs to be a mechanism to make a contribution to the platform, and for the platform to accept this contribution. This is achieved through the use of extensions and extension points, another element of the Eclipse component model.

Eclipse Architecture

Modularité et extensibilité: les clefs du succès

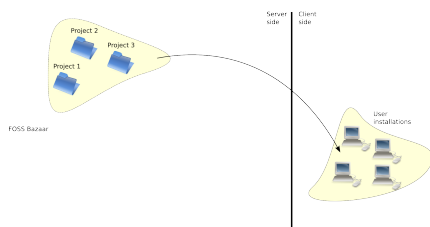
This extensible architecture was one of the keys to the successful growth of the Eclipse ecosystem. Companies or individuals could develop new plugins, and either release them as open source or sell them commercially. One of the most important concepts about Eclipse is that everything is a plugin. Whether the plugin is included in the Eclipse platform, or you write it yourself, plugins are all first class components of the assembled application.

Aujourd'hui: plus de 1000 composants sur le Eclipse Marketplace, et des milliers d'autres ailleurs!

Le cas des Distributions GNU/Linux

Le cas des distributions GNU/Linux

Avant l'arrivée des distributions, le seul moyen d'installer du logiciels sur les postes clients était:



Mais:

- ▶ pas de *version standard* du poste client
- ▶ besoin de recompiler, ... et reconfigure assez souvent
- ▶ trop compliqué!

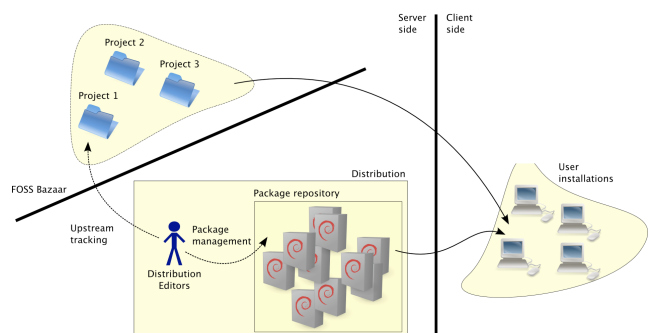
Systèmes logiciels complexes

On a vu comment l'architecture modulaire est à la base de bien de logiciels libres qui ont du succès.

Nous nous intéressons maintenant à comment on peut mettre en place une architecture modulaire pour des systèmes logiciels de grande envergure, comme les Distributions GNU/Linux.

Le cas des distributions Linux

Cela a fait naître les distributions comme intermédiaires entre les projets et les utilisateurs



Le rôle d'un éditeur de distribution:

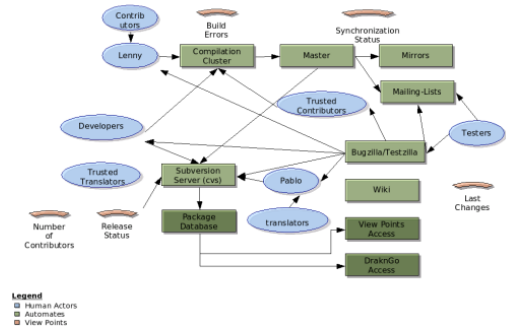
- upstream tracking** : suivre l'évolution des sources
the developer is almost never the packager!
- integration** : offrir une collection cohérente³⁹ de paquets
Pour cela, on doit traiter correctement des **dependances**
- testing** :
- distribution** : diffusion rapide sans casser l'existant

Ce n'est pas facile!

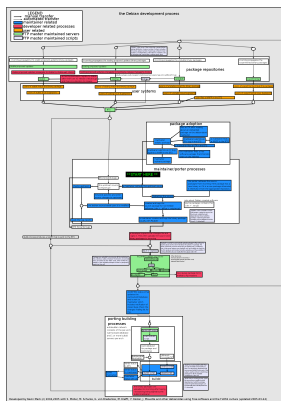
Le cycle de 6 mois de Mandriva nécessite 30 années-homme.

³⁹more formally?

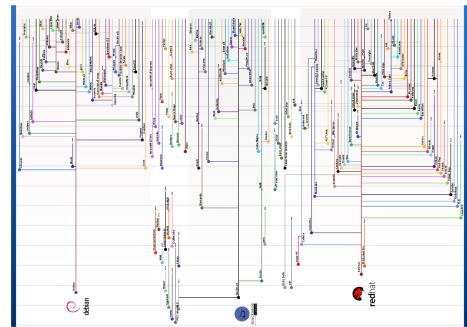
An overview of Mandriva's lifecycle



An overview of Debians's lifecycle

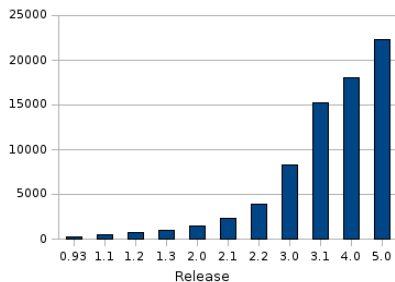


Distributions: a "somehow" successful idea . . .



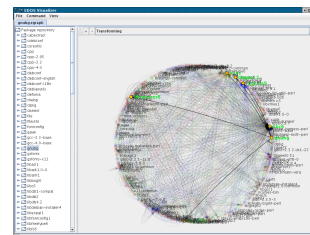
Notion centrale: pour abstraire sur la complexité de l'infrastructure existante, on utilise le **package**, avec des outils appropriés . . .

Un croissance superlinéaire



Nombre de paquets dans Debian

Des interdépendances complexes. . .



```
Package: glibc
Version: 2.14-3+20060923-4
Depends: glibc-data,
trf-bitstream-vera, libartsc0
(>= 1.5.0-1), ..., libgl1-mesa-glx
| libgl1, ...
Conflicts: ...
```

Cela change *tous les jours!* Comment s'y retrouver?

Ces problèmes sont au coeur des projets **EDOS**, et **MANCOOSI**.

Packages, metadata, installation

Package = { some files
some scripts
metadata

Example

```
Package: aterm
Version: 0.4.2-11
Section: x11
Installed-Size: 280
Maintainer: Göran Weinholt ...
Architecture: i386
Depends: libc6 (>= 2.3.2.ds1-4),
libc6-i386 (>= 4.1.0), ...
Conflicts: suidmanager (< 0.50)
Provides: x-terminal-emulator
...
```

- Identification
- Inter-package rel.
 - Dependencies
 - Conflicts
- Feature declarations
- Other
 - Package maintainer
 - Textual descriptions

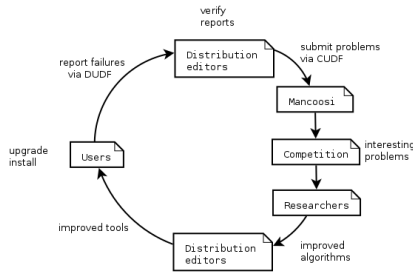
- a package is the **elemental component** of modern distribution systems (not GNU/Linux specific)
- a working **system** is deployed by installing a package set (≈ 1000/2000 for GNU/Linux distro)

Installation process

| Phase | Trace |
|-----------------------|---|
| User request | # apt-get install aterm Reading package lists... Done Building dependency tree... Done The following extra packages will be installed: libafterimage0 The following NEW packages will be installed: aterm libafterimage0 |
| Constraint resolution | 0 upgraded, 2 newly installed, 0 to remove and 1786 not upgraded. Need to get 386kB of archives. After unpacking 807kB of additional disk space will be used. Do you want to continue [Y/n]? Y |
| Package retrieval | Get: 1 http://debian.ens-cachan.fr testing/main libafterimage0 2.2.8-2 [301kB] Get: 2 http://debian.ens-cachan.fr testing/main aterm 1.0.1-4 [84.4kB] Fetched 386kB in 0s (410kB/s) |
| Pre-configuration | Selecting previously deselected package libafterimage0. (Reading database ... 294774 files and directories currently installed.) Unpacking libafterimage0 (from .../libafterimage0.2.2.8-2_1386.deb) ... Selecting previously deselected package aterm. Unpacking aterm (from .../aterm_1.0.1-4_1386.deb) ... |
| Unpacking | Setting up libafterimage0 (2.2.8-2) ... Setting up aterm (1.0.1-4) ... |
| Configuration | |

- **each** phase can fail (it actually happens quite often . . .)
- efforts should be made to identify errors as early as possible

Mancoosi: algorithmes et optimisations



La sécurité

Sécurité et Logiciel libre

- ▶ les modèles de sécurité
- ▶ l'exemple du vote électronique

Modèles de sécurité

Fermé On construit un système dont les spécifications sont secrètes.
La sécurité repose en partie sur le secret des spécifications
(Ex: Enigma).

Avantages:

- ▶ On ajoute une difficulté supplémentaire pour les casseurs

Desavantages:

- ▶ La difficulté supplémentaire donne une fausse assurance
- ▶ Modèle traditionnellement sensible à la traison
- ▶ Difficile, voir impossible à mettre en oeuvre avec des composantes disponibles sur le marché

Modèles de sécurité

Difficile, voir impossible à mettre en oeuvre avec des composantes disponibles sur le marché,...

pourquoi?

- ▶ on ne maîtrise pas le code source propriétaire
- ▶ on ne sait pas faire d'audit de sécurité complet automatique sur 50M lignes de code
- ▶ chacune de ces lignes de code est suspecte
- ▶ même les outils de compilation⁴⁰ sont suspects!

⁴⁰Ken Thompson: Reflections on trusting trust

Reflections on trusting trust

Ken Thompson, Turing Award, 1983.

Voir l'explication au tableau.

The moral is obvious. You can't trust code that you did not totally create yourself. (Especially code from companies that employ people like me.) No amount of source-level verification or scrutiny will protect you from using untrusted code. In demonstrating the possibility of this kind of attack, I picked on the C compiler. I could have picked on any program-handling program such as an assembler, a loader, or even hardware microcode.

Communication of the ACM, Vol. 27, No. 8, August 1984, pp. 761-763.

Modèles de sécurité

Ouvert les spécifications du système sont connues de tous.
La sécurité repose aussi sur la vérification par les paires (Ex: RSA).

Avantages:

- ▶ Modèle étanche à la traison (inutile de voler un decodeur RSA).
- ▶ Modèle facile à mettre en place avec des composantes du marché, à condition qu'elles soient aussi "ouvertes" (i.e. livrées avec leur code, et ... vérifiés et reconstruits par vous!)

Desavantages:

- ▶ La qualité de la protection depend seulement de la qualité des algorithmes et de leur mise en oeuvre (crypto encore une art).
- ▶ La qualité de la protection depende aussi de la rapidité de correction des erreurs de conception ou mise en oeuvre (necessite encore du logiciel libre)

Microsoft Security Track Record

Toute information *peut* être *recueillie*, *tracée* et *transmise à votre insu*.

Microsoft l'a fait, le fait et continuera à le faire pour plusieurs raisons:

lutte aux copies illégales :

- ▶ Windows Registration Wizard (<http://www.1emonde.fr/nvtechno/gates/pirate.html>)
- ▶ Licence Windows Media Player
- ▶ fichiers Office avec GUID
- ▶ projet Tempest

WMP EULA (version 8, 2002)

** Digital Rights Management (Security). You agree that in order to protect the integrity of content and software protected by digital rights management ("Secure Content"), Microsoft may provide security related updates to the OS Components that will be automatically downloaded onto your computer. These security related updates may disable your ability to copy and/or play Secure Content and use other software on your computer. If we provide such a security update, we will use reasonable efforts to post notices on a web site explaining the update.*

Un exemple: le vote électronique

Comme quoi le code source ne suffit pas...

Le vote électronique: un exemple de l'existant

Extrait de http://www.alexandriavoter.org/eSlate/eSlate_slide_show.html:



Le vote électronique: un exemple de l'existant

Extrait de http://www.alexandriavoter.org/eSlate/eSlate_slide_show.html:



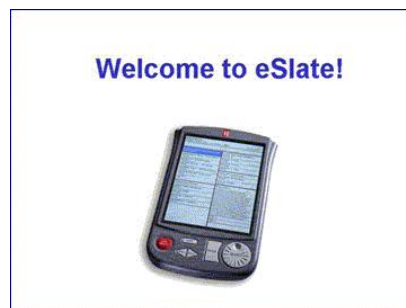
Mais aujourd'hui, c'est bien pire...

- ▶ AppStore totalitaire
- ▶ filtrage du web, et collecte d'informations personnelles
- ▶ publicité qui vous suit
- ▶ flux d'information entre sites web
- ▶ deep-packet inspection...

Les politiques se reveillent (un peu tard):
<http://privacybydesign.ca/>

Le vote électronique: un exemple de l'existant

Extrait de http://www.alexandriavoter.org/eSlate/eSlate_slide_show.html:



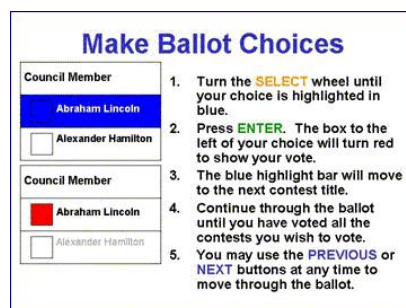
Le vote électronique: un exemple de l'existant

Extrait de http://www.alexandriavoter.org/eSlate/eSlate_slide_show.html:



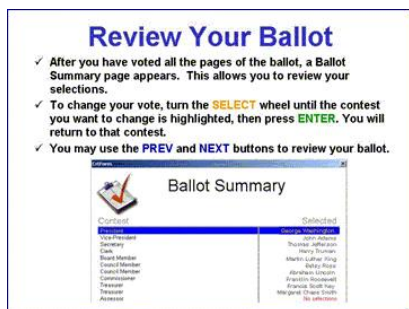
Le vote électronique: un exemple de l'existant

Extrait de http://www.alexandriavoter.org/eSlate/eSlate_slide_show.html:



Le vote électronique: un exemple de l'existant

Extrait de http://www.alexandriavoter.org/eSlate/eSlate_slide_show.html:



Le vote électronique: un exemple de l'existant

Extrait de http://www.alexandriavoter.org/eSlate/eSlate_slide_show.html:



Le vote électronique: un exemple de l'existant

Extrait de http://www.alexandriavoter.org/eSlate/eSlate_slide_show.html:



Est-ce que vous achetez ça?
Quel modèle de "sécurité" peut nous donner confiance dans des tels systèmes?

Logiciels Libres et vote électronique

Certains disent que:

Le code source des machines de vote électronique doit être entièrement du logiciel libre.

Est-ce suffisant? Voir

<http://avirubin.com/vote/response.html> pour une analyse du code source (non libre!) des machines Diebold.

Les propriétés d'un protocole de vote électronique

- pas de bourrage des urnes I seulement les électeurs peuvent voter
 - pas de bourrage des urnes II personne peut voter plus d'une fois
 - pas de bourrage des urnes III personne peut voter à la place de quelqu'un d'autre
 - anonymat du vote personne connaît ce que *quelqu'un d'autre* a voté
 - contrôle cela peut prendre deux formes:
 - ▶ chaque électeur peut vérifier que son vote est bien pris correctement en compte
 - ▶ chacun peut vérifier que le vote de chaque électeur est bien pris correctement en compte
 - pas de coercition personne ne peut "prouver" d'avoir voté dans un sens ou dans l'autre
- Notez bien que ces deux dernières propriétés semblent incompatibles!

Une analyse du problème

- ▶ le vote à préférences multiples ordonnées
- ▶ les protocoles de vote électronique

Comparer avec l'existant

Discussion ouverte sur les modèles de développement du logiciel