HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Master's Thesis: Censorship-resistant Collaboration with a Hybrid DTN/P2P Network

Philipp Hagemeister

Institut für Informatik
Heinrich-Heine-Universität Düsseldorf

29.3.2012

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Threat Model

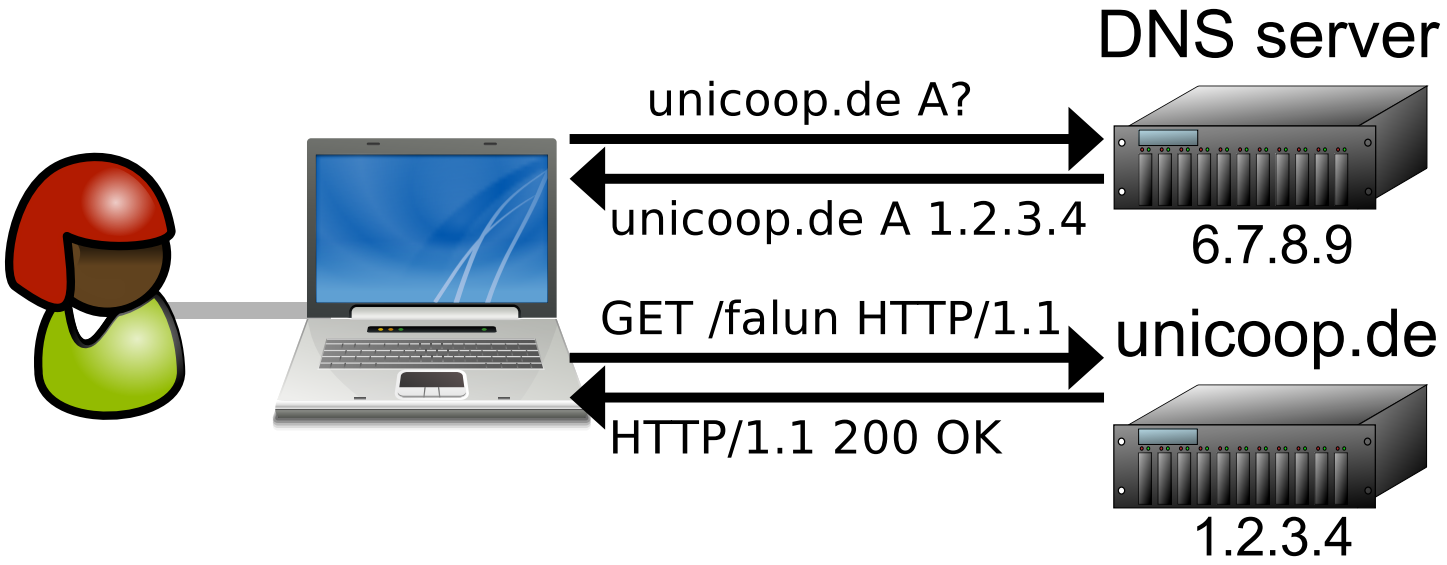
- Collaboration systems proliferate free speech
- Attacker does not want free speech
- $\Rightarrow$ Attacker goal: Disrupt collaboration systems
- Attacker controls ISP and national infrastructure

Figure: Attacker (representation)

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Requests in Current Collaboration Systems

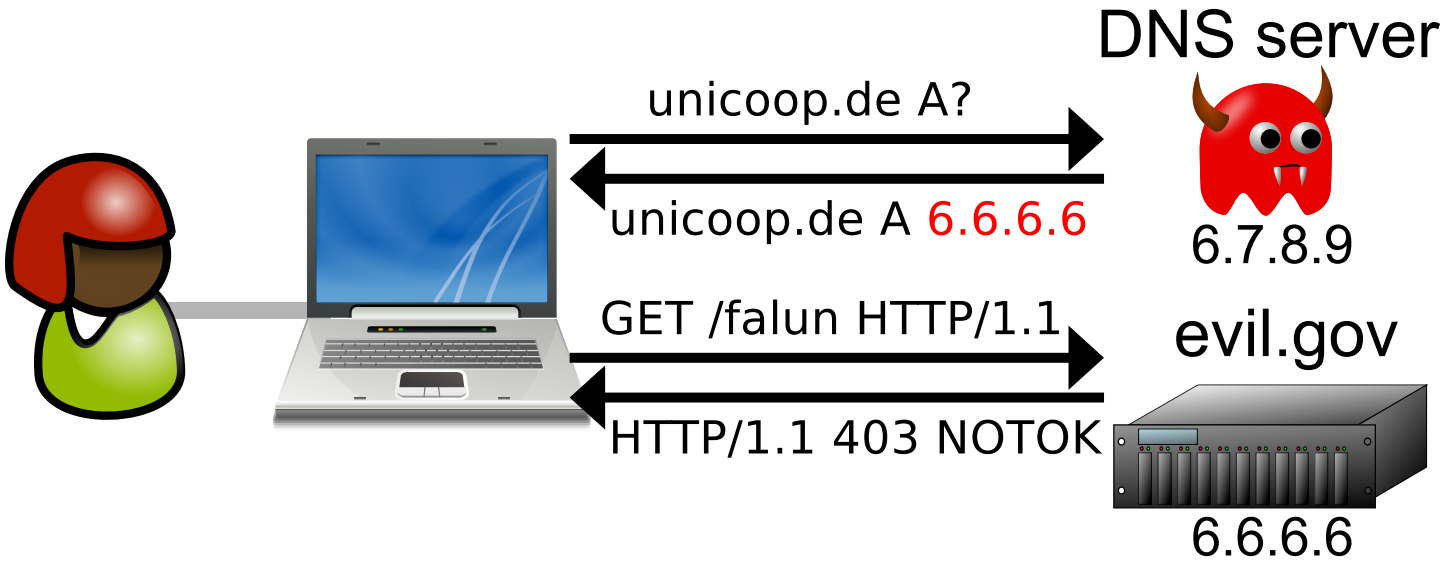
- adhocracy, echo, LiquidFeedback, UniCoop are web applications
- Request diagram:



HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

DNS Censorship

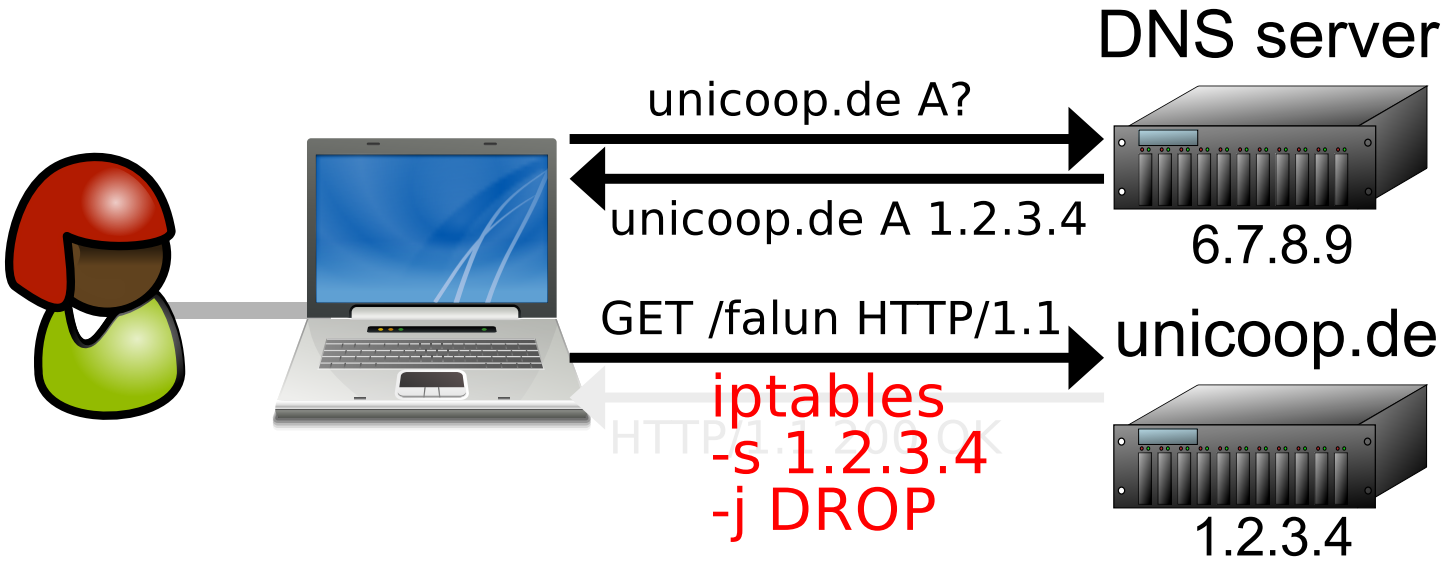
- Attacker controls default DNS server
- Contemplated in Germany and US
- Used in Belgium, Denmark, Italy, Turkey, Burma, China, ...
- Easily circumvented (→ Allessandro Lenzen, 2011)
- Long-term solution: client-side DNSSec



HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

IP Censorship

- Attacker can drops packets from or to specific IP addresses
- Used in China, Egypt, Libya, Pakistan, Thailand



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

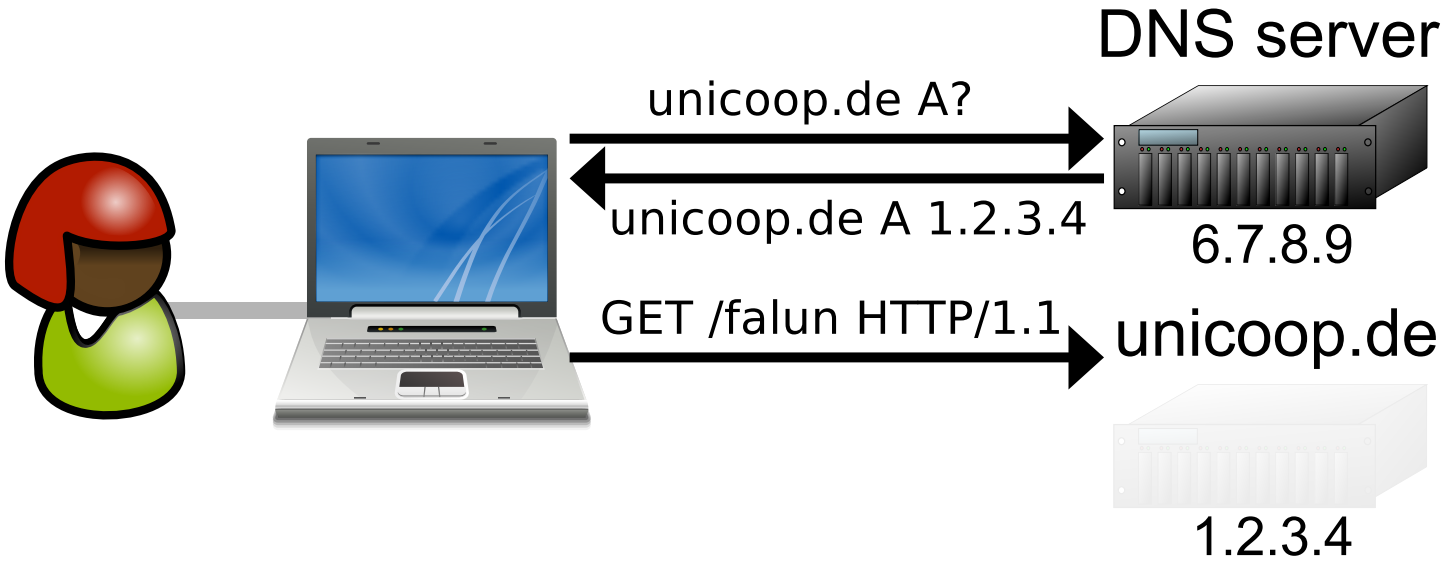
Deep Packet Inspection

- Attacker filters packets for search terms
- Used in China, Iran
- Prevented by encryption

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

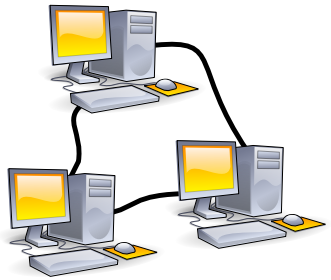
Physical Attacks

- Attacker physically seizes or takes over server
- Happened in Germany!
  - In 2011, servers of the Piratenpartei were confiscated
- Defense: Multiple servers



Peer-To-Peer (P2P) Networks

- Multiple servers alone are not sufficient
- Eliminate all single points of failure!
- We need a truly decentralized system
- ... a **Peer-to-peer** (**P2P**) network



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Bootstrapping

How do we get the address of a peer?

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Bootstrapping

How do we get the address of a peer?

- Hardcoded
- Human input

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Bootstrapping

How do we get the address of a peer?

- Hardcoded
- Human input
- DNS
- HTTP(S)

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Bootstrapping

How do we get the address of a peer?

- Hardcoded
- Human input
- DNS
- HTTP(S)
- IP multicast

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Bootstrapping

How do we get the address of a peer?

- Hardcoded
- Human input
- DNS
- HTTP(S)
- IP multicast
- Email / SMS

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Bootstrapping

How do we get the address of a peer?

- Hardcoded
- Human input
- DNS
- HTTP(S)
- IP multicast
- Email / SMS
- Decoy routing

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Bootstrapping

How do we get the address of a peer?

- Hardcoded
- Human input
- DNS
- HTTP(S)
- IP multicast
- Email / SMS
- Decoy routing

Bootstrapping: Solvable

The number of bootstrapping schemes allow us to evade all but the most sophisticated censorship systems.

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Other P2P considerations

- Structured vs unstructured
- Sybil and other active attacks
- Broadcasting
- NAT traversal

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Other P2P considerations

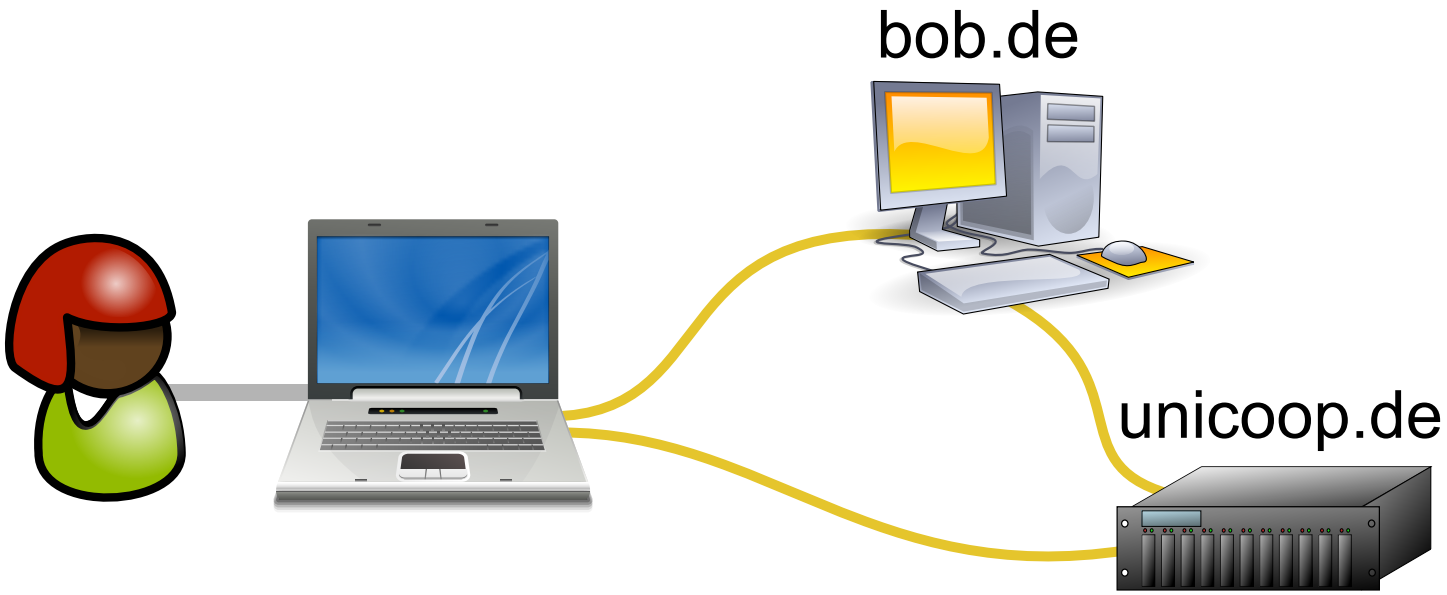
- Structured vs unstructured
- Sybil and other active attacks
- Broadcasting
- NAT traversal
- Privacy
  - Solved by **anonymization networks**
  - Examples: I2P, Tor, Freenet
  - Need to be integrated
  - → Paul Baade

P2P: Conclusion

A P2P network can provide an adequate defense against censorship.

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Back to the Threat Model



HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Total Internet Shutoff

- Attacker can turn off Internet access
- Happened in 2011 in Egypt and Libya
- Arguably permanently in Cuba and North Korea

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

DTN

- Transfer data with USB thumb drives
- Delay-Tolerant Networks (**DTN**s) do not require continuous connection
- Fields of use:
  - Interplanetary communication
  - Developing nations
  - Military/naval
  - Sneakernet in Cuba

DTNs allow communication even in the case of a Internet shutoff

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Revision Control in DTNs

- Challenge in DTNs: Distributed consensus is not possible
- Nevertheless, we want want **revision control**
  - .. primarily for history, accountability, and change management

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Revision Control in DTNs

- Challenge in DTNs: Distributed consensus is not possible
- Nevertheless, we want want **revision control**
  - .. primarily for history, accountability, and change management
- Graph-based revision control systems: git, mercurial, bazaar, PlatinVC
  - Need to be adapted for DTNs (→ Janine Haas, 2012)

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Revision Control in DTNs

- Challenge in DTNs: Distributed consensus is not possible
- Nevertheless, we want want **revision control**
  - .. primarily for history, accountability, and change management
- Graph-based revision control systems: git, mercurial, bazaar, PlatinVC
  - Need to be adapted for DTNs (→ Janine Haas, 2012)
- Patch-based revision control systems: darcs
  - Complex, not yet practical
  - Could be the silver bullet

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Revision Control in DTNs

- Challenge in DTNs: Distributed consensus is not possible
- Nevertheless, we want want **revision control**
  - .. primarily for history, accountability, and change management
- Graph-based revision control systems: git, mercurial, bazaar, PlatinVC
  - Need to be adapted for DTNs (→ Janine Haas, 2012)
- Patch-based revision control systems: darcs
  - Complex, not yet practical
  - Could be the silver bullet
- Document-oriented revision control: CouchDB, MongoDB
  - Simple, but weak guarantees

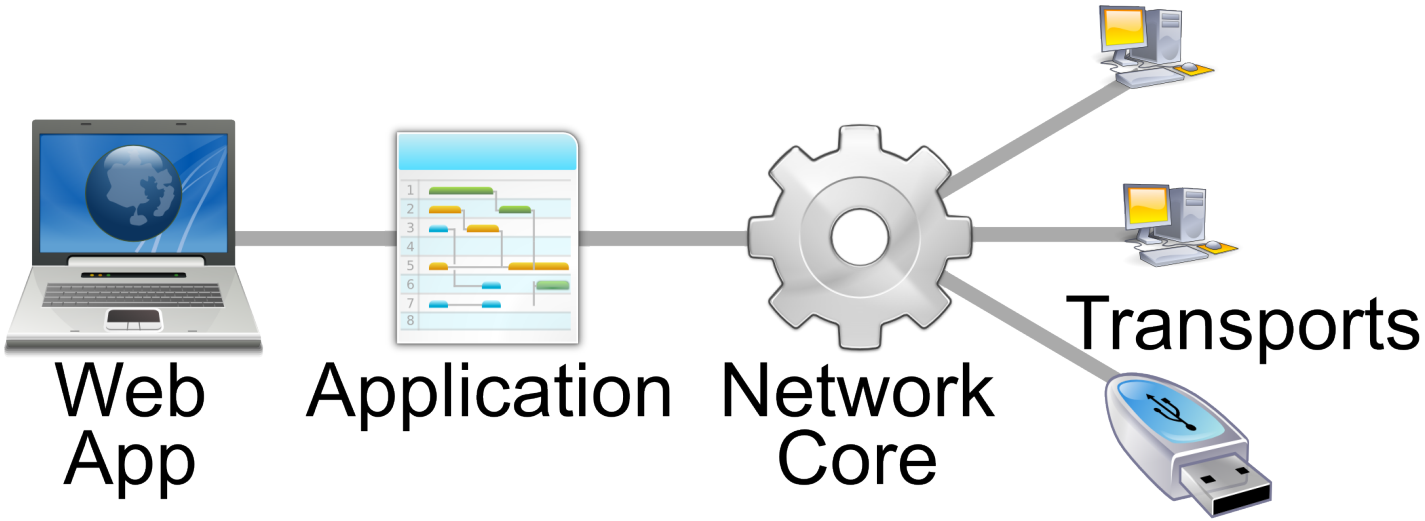
HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Revision Control in DTNs

- Challenge in DTNs: Distributed consensus is not possible
- Nevertheless, we want want **revision control**
  - .. primarily for history, accountability, and change management
- Graph-based revision control systems: git, mercurial, bazaar, PlatinVC
  - Need to be adapted for DTNs ($\rightarrow$ Janine Haas, 2012)
- Patch-based revision control systems: darcs
  - Complex, not yet practical
  - Could be the silver bullet
- Document-oriented revision control: CouchDB, MongoDB
  - Simple, but weak guarantees
- Common base technology: **Content-Adressable Storage**(**CAS**)
  - Stores a set of bytes, accessed with hash(bytes).
  - No conflicts, $sync(CAS1, CAS2) = CAS1 \cup CAS2$
  - Can store (almost) all of the revision control system data

Architecture

- *Transports* abstract the specific communication channel
  - P2P over TCP
  - DTN over USB thumb drive
  - P2P over anonymization network
  - DTN over facebook
- Requirement: Application should be available on every device
  $\Rightarrow$ web application



HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Web Application Fallback

- Best experience (DTN) if system is locally installed
- Public web servers for the masses
- If a web server becomes unavailable, switch to another one

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Web Application Fallback

- Best experience (DTN) if system is locally installed
- Public web servers for the masses
- If a web server becomes unavailable, switch to another one
- Alternative: Continue working offline (→ Tim van Cleef)
- Future: Whole application in the browser

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Conclusion

- Censorship resistance is important for collaboration software
- Censorship-resistant P2P network
- In case of total shutoff: DTN
- Future reasearch and implementation required

Questions?

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Questions?

This presentation: `http://phihag.de/2012/mtpres.pdf`
Thesis: `http://phihag.de/2012/mt.pdf`
Source code: `http://phihag.de/2012/d2p/`

Demo

Warning: Experimental Prototype!

Future Work

- General code quality, documentation, and testing
- Automated unit and functional tests
- Simulation framework
- P2P bootstrap implementation and analysis
- NAT traversal for the P2P transport
- Structured P2P implementation with efficient broadcast
- Integration into DTN standards (RFC 4838 ...)
- Research into partial replication
- Robust thumb drive storage formats
- Steganography and cryptography
- Ports to other platforms, in particular android, *BSD, iOS, Mac OS X, WebOS, Windows, Windows Phone

Future Work (continued)

- Project search functionality
- User Management
- Extend functionality of the main policy drafting application
  - A WYSIWYG editor
  - Comments to specific lines or paragraphs (→ Julius Römmler)
  - Better usability
- Demonstrate and develop a client-side application
- Prototype browser-to-browser P2P with WebRTC
- Create a decentralized security framework
- Allow closed groups as well as read-only ones
- Allow voting applications
- Extend revision control
  - Integrate graph- and/or patch-based revision control systems
  - Improve the CAS performance
- Integration with other platforms (such as adhocracy)
- Integration with PKIs such as German ID card

Security

- Problem: Where do we store keys
- Browser integration problematic (→ Evgeni Golov, 2012)
- Option: private key = hash(password)

Security

- Problem: Where do we store keys
- Browser integration problematic ($\rightarrow$ Evgeni Golov, 2012)
- Option: private key = hash(password)
- $\forall$ project:
  - Project ID = hash(project public key, security specification)
  - Allow private projects by encrypting everything with a symmetric key
  - Symetric key is stored alongside project data, encrypted with users' public keys

Security

- Problem: Where do we store keys
- Browser integration problematic (→ Evgeni Golov, 2012)
- Option: private key = hash(password)
- ∀ project:
  - Project ID = hash(project public key, security specification)
  - Allow private projects by encrypting everything with a symmetric key
  - Symetric key is stored alongside project data, encrypted with users' public keys
  - Allow read-only projects by requiring changes to be signed by a key ...
  - ... which in turn is signed by the project's key

Voting

- Distributed verifiable anonymous voting is not possible!
- Requires trusted intermediaries
- Or trusted voting registrars

Extended Threat Model

- Assumption so far: *User can run arbitrary software on her device.*
- Assumption: User *has access to* a device
- Assumption: User controls (general-purpose) device.
  - May be restricted with UEFI Secure Boot
  - Signed firmware required on Apple iPad, iPhone, iPod
  - Signed firmware required on some android devices
- Attacker may also physically go after users
  - $\Rightarrow$ Anonymity/Pseudonimity required

Extended Threat Model

- Assumption so far: *User can run arbitrary software on her device.*
- Assumption: User *has access to* a device
- Assumption: User controls (general-purpose) device.
  - May be restricted with UEFI Secure Boot
  - Signed firmware required on Apple iPad, iPhone, iPod
  - Signed firmware required on some android devices
- Attacker may also physically go after users
  - $\Rightarrow$ Anonymity/Pseudonimity required

- Attacker can use **malware** to gain control of the device
  - Happened in Germany: Staatstrojaner
  - Blackberry malware in UAE

DPI in China

- Chinese network-level DPI searches for keywords like *falun gong*
- Injects an `RST` packet
- Blocks all packets between the peers for a couple of minutes

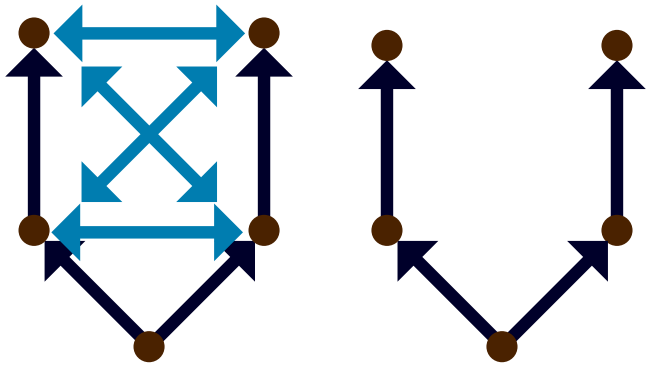
192.168.1.13	192.168.1.1	DNS	70	Standard query A pku.edu.cn
192.168.1.13	192.168.1.1	DNS	70	Standard query AAAA pku.edu.cn
192.168.1.1	192.168.1.13	DNS	120	Standard query response
192.168.1.1	192.168.1.13	DNS	102	Standard query response A 162.105.129.21 A 162.10
192.168.1.13	162.105.129.21	TCP	74	56558 > http [SYN] Seq=0 Win=14600 Len=0 MSS=1460
162.105.129.21	192.168.1.13	TCP	58	http > 56558 [SYN, ACK] Seq=0 Ack=1 Win=3840 Len=
192.168.1.13	162.105.129.21	TCP	54	56558 > http [ACK] Seq=1 Ack=1 Win=14600 Len=0
192.168.1.13	162.105.129.21	HTTP	128	HEAD /faluX_gXng HTTP/1.1
162.105.129.21	192.168.1.13	TCP	54	http > 56558 [ACK] Seq=1 Ack=75 Win=5840 Len=0
162.105.129.21	192.168.1.13	TCP	259	[TCP segment of a reassembled PDU]
192.168.1.13	162.105.129.21	TCP	54	56558 > http [ACK] Seq=75 Ack=206 Win=15544 Len=0
192.168.1.13	162.105.129.21	HTTP	128	HEAD /falun_gong HTTP/1.1
162.105.129.21	192.168.1.13	TCP	54	http > 56558 [RST, ACK] Seq=206 Ack=149 Win=1923

P2P: Structured vs Unstructured

- Structured networks are stable
- But may be easier to disrupt!

P2P: Structured vs Unstructured

- Structured networks are stable
- But may be easier to disrupt!
- Broadcasting much more efficient in structured networks

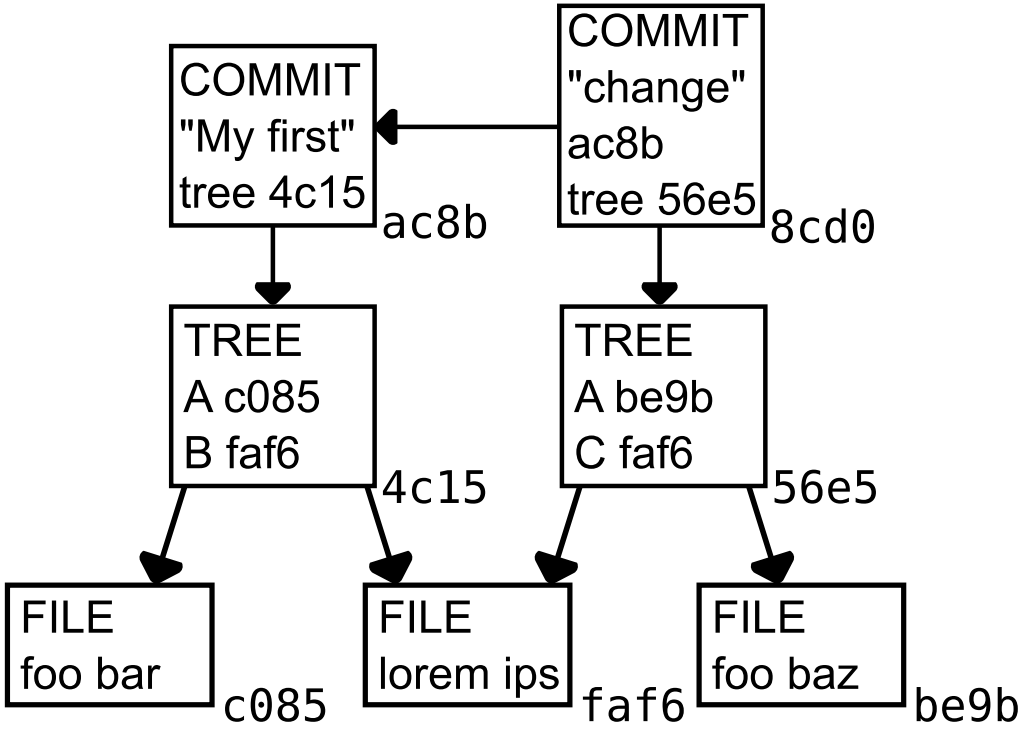


Unstructured
12 Transmissions

Structured
4 Transmissions

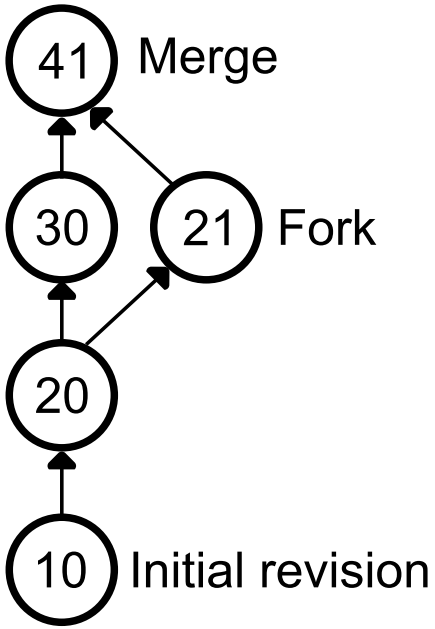
Graph-based Revision Control Systems

- Every file, tree, commit is mapped to a block of content
- Block is stored in a CAS
- Accessible only by `hash(block)`



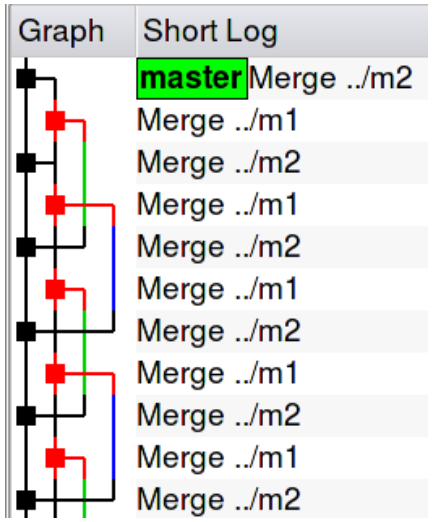
Terms in Revision Control Systems

- Every change is recorded in a *commit*
- Commits form a *DAG*:



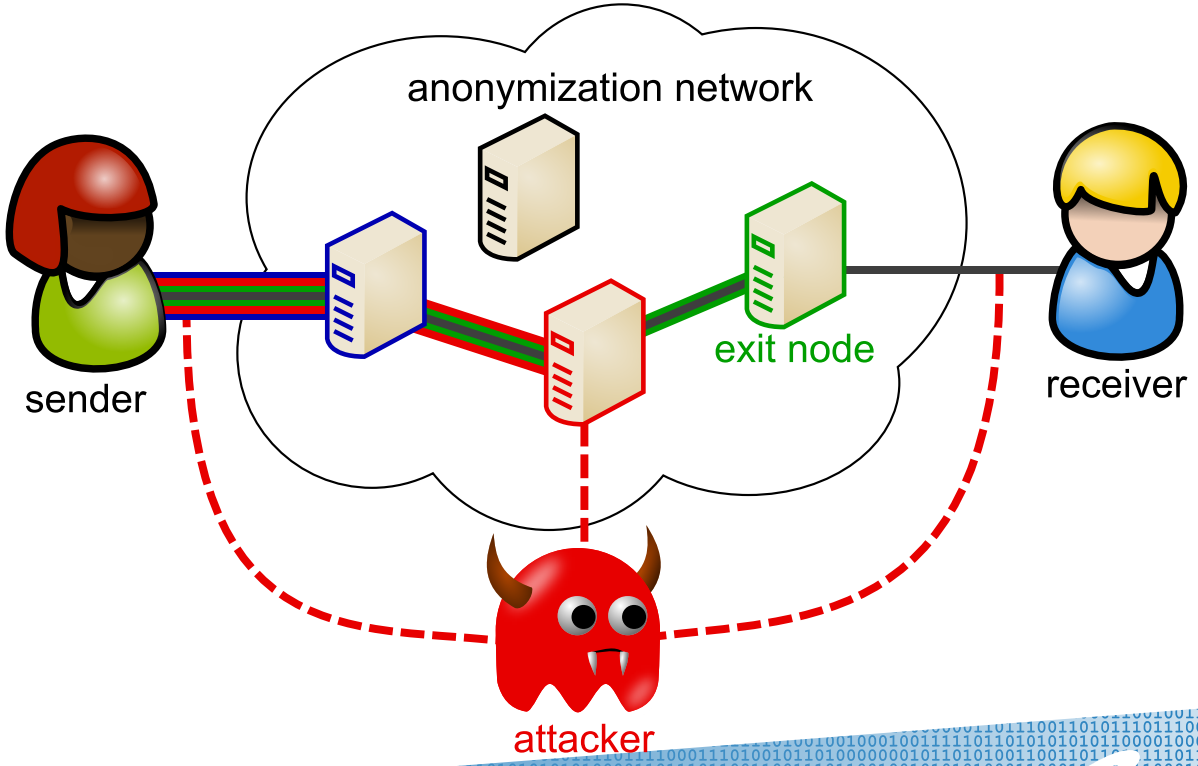
Problems in Graph-based Revision Control Systems

- Assumption: Always one common HEAD
- Problem: **D**elays mean that automatic merging can go on forever



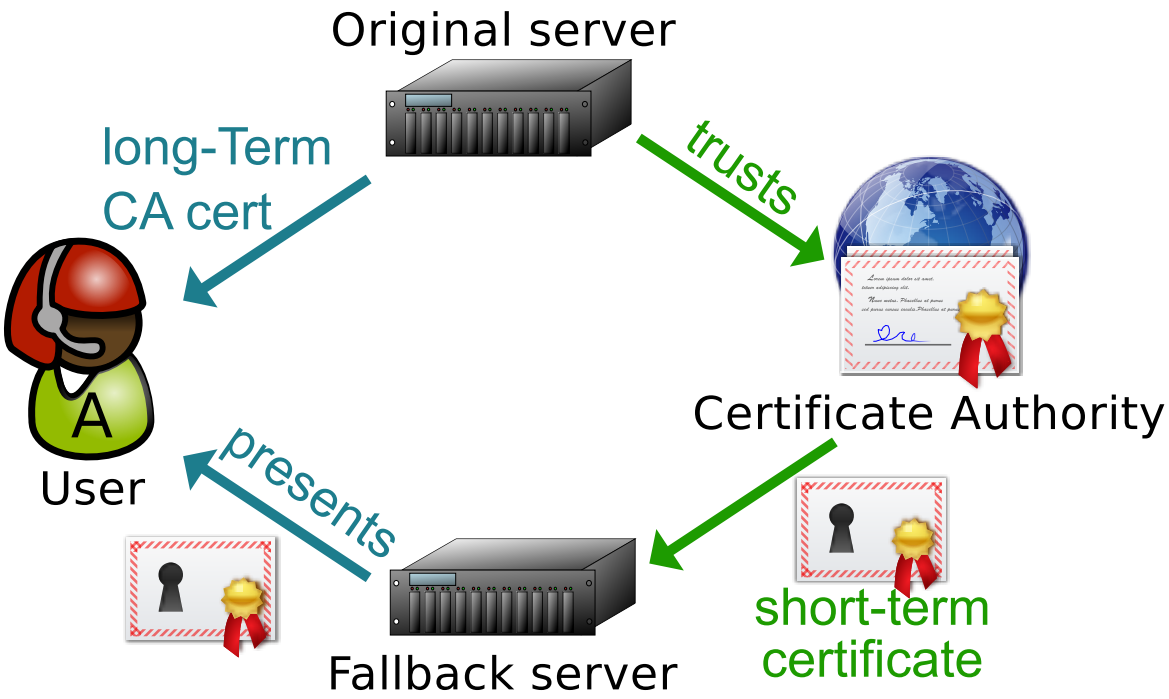
Anonymization Frameworks

- Use a user-chosen combination of *mixes*
- Tor (bidirectional, TCP-like)
- I2P (unidirectional, UDP-like)
- GnuNet (only storage)



Web Fallback Verification

- Problem: What if attacker compromises a server?
- Solution: Short-term certificates
- CA(might be blocked) does never interact with user



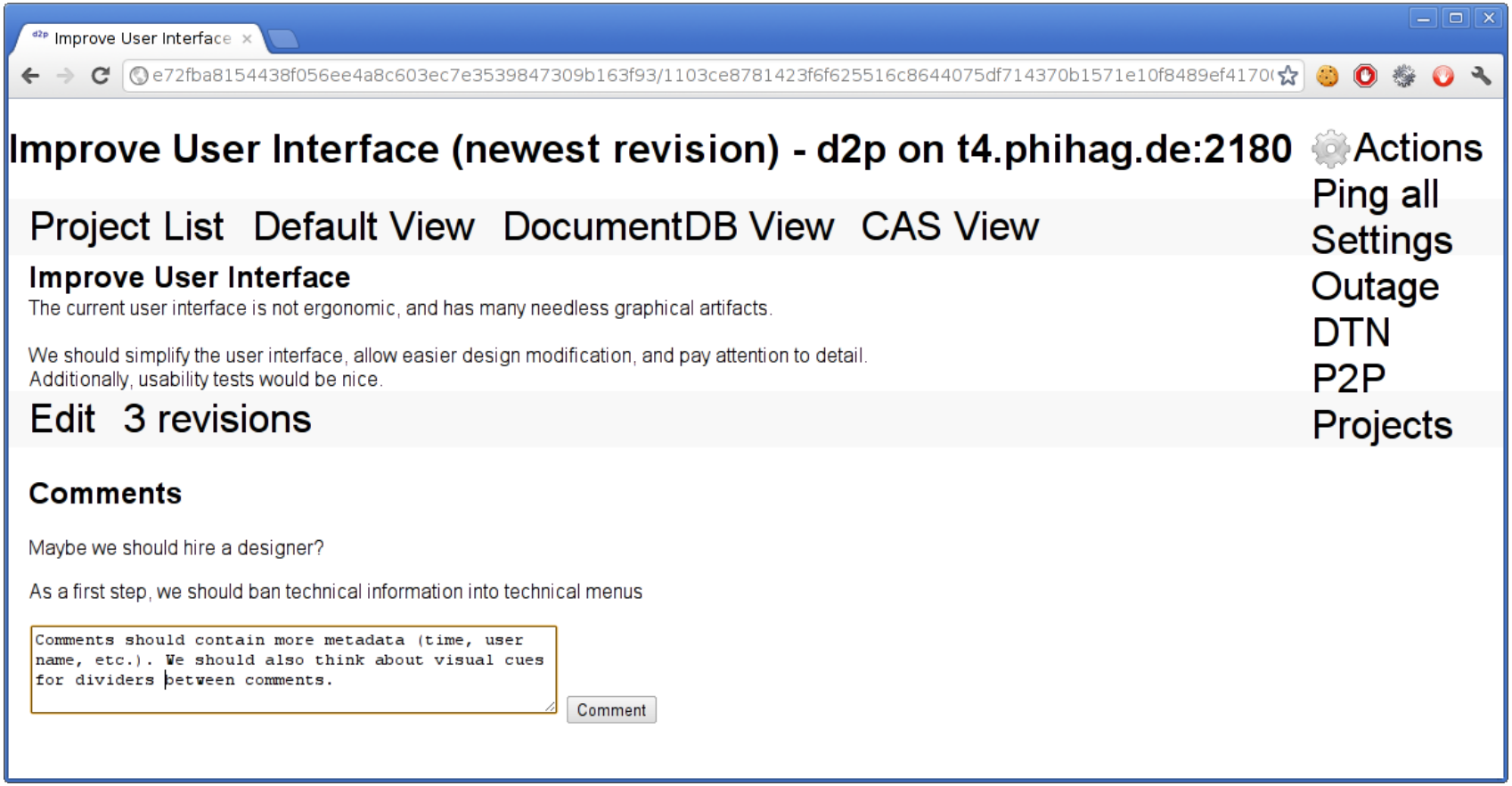
Implementation Considerations

- Code (especially views) must be **portable**
  - Required for offline version (→ Tim van Cleef)
  - We may also want to reimplement/compiler the application for the browser
  - Mustache: Logic-less web templates

Implementation Considerations

- Code (especially views) must be **portable**
  - Required for offline version (→ Tim van Cleef)
  - We may also want to reimplement/compiler the application for the browser
  - Mustache: Logic-less web templates
- Python **3** for clean code (Why not 2? bytes vs `str`ing)
- **Tornado** as asynchronous framework
- Modern web technologies (WebSocket, WebRTC, HTML5 semantic elements)
- Automated tests, simulation

Screenshots (1)



Screenshots (2)