Towards Trustworthy Online Voting: Distributed Aggregation of Confidential Data

PhD Thesis Defense

Robert Riemann

Inria/ENS de Lyon

18th December 2017

Trust in Authorities

Fig. Nuclear Power
© Flickr/bagalute
(CC by)

Fig. Drugs
© Flickr/F. E. A. Nisar
(public domain)

Fig. Food
© Flickr/kgregory
(CC by-nc-nd)

Without consensus on trusted authorities, it is reasonable to omit authorities altogether.

Compare development to:

- **Bitcoin**[1]
  gold, fiat money, online banks, Bitcoin
- **BitTorrent**[2]
  circulating disks, FTP (web server), BitTorrent

[1] S. Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008.
[2] B. Cohen. The BitTorrent Protocol Specification. 2008.

LES PROJETS LAURÉATS 2017

Tout Paris > Par arrondissement >

LES LAURÉATS POUR TOUT PARIS

PROPRETÉ

TOUT PARIS

15 - #VillePlusPropre

3 000 000 €

ENVIRONNEMENT

TOUT PARIS

13 - #SousLesPavésDesF

3 500 000 €

SOLIDARITÉ ET COHÉSION SOCIALE

TOUT PARIS

20 - #VilleRefuge

5 000 000 €

Outline

Complex Cooperation

Online Services are among the largest cooperations.
Facebook counts 2 billion monthly active users.

Online Service emerge in all areas of life:

- Commerce (Alibaba, Amazon)
- Social Networks (Facebook, Twitter, Weibo, VK)
- Intermediary Services (AirBnB, Uber)
- eGovernment (Registries, Taxation, eParticipation)

Common Observation: governed by operators (authorities)

Promoting Trust

1. How to ensure trust in cooperation?
2. How to govern large cooperations?

Good Governance Principles promoting Trust[3]

Characteristics beneficial for Trust:

- Transparency
- Participation
- Accountability

Characteristics beneficial for Scalability:

- Responsiveness
- Efficiency

[3] UNESCAP. “What is Good Governance ?”. In: United Nations Economic and social Comission for Asia and the Pacific (2009).

Good Governance Principles promoting Trust[3]

Characteristics beneficial for Trust:

- Transparency
- Participation
- Accountability

Characteristics beneficial for Scalability:

- Responsiveness
- Efficiency, which includes somehow
- Convenience

[3] UNESCAP. "What is Good Governance ?". In: United Nations Economic and social Comission for Asia and the Pacific (2009).

Trust in Cooperation

Personal Trust

- based on personal relationships among cooperation members

Institutional Trust

- based on organisational security
- e.g. division of power and checks and balances

Technological Trust

- based on physical security
- e.g. barriers, locks and cryptography

become a supporter subscribe find a job

theguardian

news sport opinion arts lifestyle

UK world business football environment tech UK politics science

iOS

Apple fixes HomeKit bug that allowed remote unlocking of users' doors

Security flaw in latest iPhone and iPad iOS 11.2 software meant hackers could potentially gain remote control of lights, cameras and locks in smart homes

Fig. Guardian Article by S. Gibbs published on 8th December 2017

Technology Impact on Voting

Fig. Digital Natives.
© Flickr/antmcneill (CC by-sa)

Fig. Paper-based Voting.
© Flickr/coventrycc (CC by-nd-nd)

Classical Online Voting Security Concepts

- **Trusted Authorities**
  essentially give up secrecy and correctness
- **Anonymous Voting**
  assume unlinkability of distinct communication channels
- **Random Pertubation**
  assume shuffle of encrypted votes before their decryption
- **Homomorphic Encryption**
  assume aggregation of encrypted votes before decryption

Identified Issues

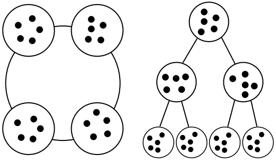
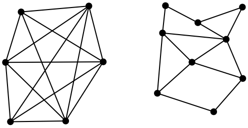
- concentration of power
- concentration of data

From Centralised to Distributed Online Voting

What if all voters become authorities?

- reuse existing protocols with:
  distributed key generation and threshold decryption
- fits the purpose of small board room votings
- does not scale

Review of Distributed Online Voting

(a) DPol (b) SPP (c) SMC (d) Blockchain

- **Secure Multi-party Computation (SMC)**
  communication in $\mathcal{O}\left(n^2\right)$, for board room votings
- **Distributed Polling (DPol)**
  secret sharing scheme applied to groups aligned in a circle
- **Secure and Private Polling (SPP)**
  SMC and threshold decryption applied to groups in a tree
- **Blockchain-based Voting**
  Bitcoin to aggregate votes (coloured coins)

Taxonomy of Distributed Online Voting[7]

Protocol	Degree of Special.	Topology	Distrib. Phases
Paper-based	none (flexible)	distributed	all
Helios,[4]	selected authorities	centralised	verification
DPol,[5]	none	structured, ring	all
SPP,[6]	random authorities	structured, tree	aggregation
Blockchain-based	none (flexible)	distributed	all

[4] B. Adida. "Helios: Web-based Open-Audit Voting.". In: USENIX Security Symposium 17 (2008), pp. 335–348.

[5] R. Guerraoui et al. "Decentralized polling with respectable participants". In: Journal of Parallel and Distributed Computing 72.1 (Jan. 2012), pp. 13–26.

[6] S. Gambs et al. "Scalable and Secure Aggregation in Distrib. Networks". In: IEEE 31. Symp. on Reliable Distributed Systems. 2011, pp. 181–190.

[7] R. Riemann and S. Grumbach. "Distributed Protocols at the Rescue for Trustworthy Online Voting". In: Proc. of the 3rd Int. Conf. on Information Systems Security and Privacy (ICISSP). Porto, Feb. 2017.

Taxonomy of Distributed Online Voting

Protocol	Degree of Special.	Topology	Distrib. Phases
Paper-based	none (flexible)	distributed	all
Helios	selected authorities	centralised	verification
DPol	none	structured, ring	all
SPP	random authorities	structured, tree	aggregation
Blockchain-based	none (flexible)	distributed	all

Remarks:

- Blockchain-based protocols are most promising for their similarity with paper-based voting
- To our knowledge: no publication yet on scalable Blockchain-based protocols

BitBallot

BitBallot[8] is a P2P aggregation protocol for online voting.

Principle Concepts:

- Pull Principle (pull gossiping to spread information)
- Aggregation over a Tree (peers assigned to leaves)
- Aggregation as a Middleware

Aggregation Operation

$\oplus : \mathbb{A} \times \mathbb{A} \mapsto \mathbb{A}$ with $\oplus$ commutative and associative

[8] D. Reimert et al. "Machine de Vote électronique et Infrastructure comportant une telle Machine". Patent FR 3037702 (France). Dec. 23, 2016.

BitBallot: Aggregation

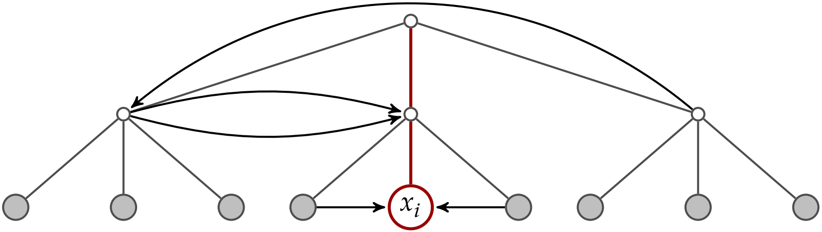


Fig. Exemplary flow of information to a peer P_i with leaf node x_i according to the pull principle of BitBallot on top of a tree overlay. Peers (in gray) respond to pull calls from P_i. Intermediate tree nodes represent any peer of the respective subtree.

BitBallot: Scalability

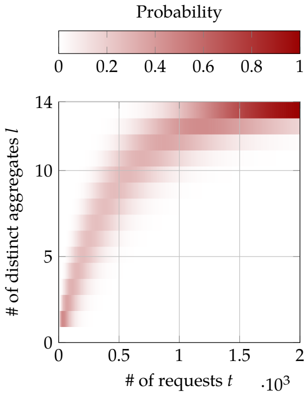


Fig. Probability $p_{r,l}$ that a peer has l distinct foreign aggregates after t requests. A tree with arity $k = 15$ and depth $d = 2$ is considered. P_i joins the aggregation when all 14 sibling peers have already acquired their 14 aggregates.

Conclusion:

- P_i can reconstruct parts of the tree from given responses
- obfuscation of source leads to significant overhead

ADVOKAT

ADVOKAT[9][10] is a new P2P aggregation protocol.

Principle Concepts:

- Peer Discovery and Routing based on Kademlia
- Aggregation over Binary Tree (of Kademlia)
- Distributed Tree Configuration
- Extensions to improve Correctness based on Signatures

[9] Aggregation for distributed voting online using the Kademlia DHT

[10] R. Riemann and S. Grumbach. "Secure and trustable distributed aggregation based on Kademlia". In: IFIP Advances in Information and Communication Technology. Vol. 502. Rome, May 2017. Chap. 12.

Distributed Hash Table Kademlia[11] for Routing

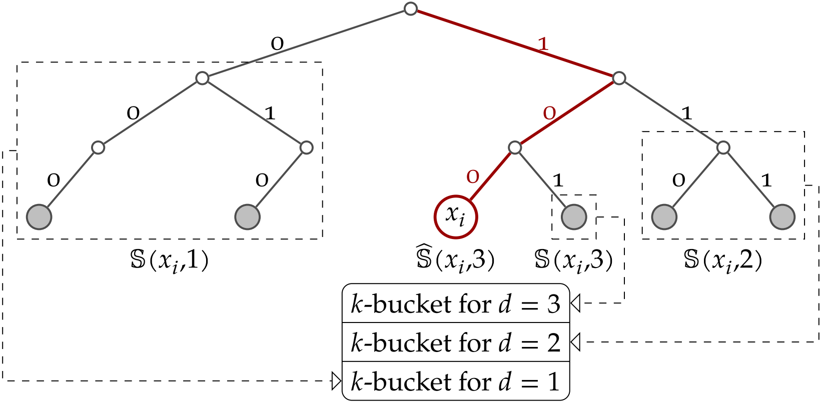
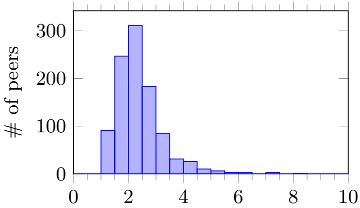


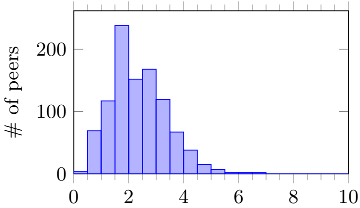
Fig. Kademlia Tree

[11] P. Maymounkov and D. Mazieres. "Kademlia: A peer-to-peer information system based on the xor metric". In: 1st Int. Workshop on P2P Systems (2002), pp. 53–65. DOI: 10.1007/3-540-45748-8_5.

Confidentiality: Knowledge Distribution



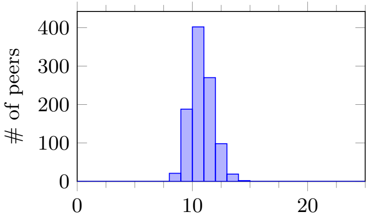
(a) Histogram of leaked information L_i.



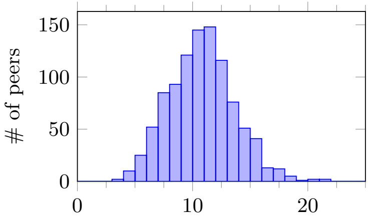
(b) Histogram of received information R_i.

In a simulation with $n = 1000$, peers leak (a), respectively receive (b), information on initial aggregates depending on the global distribution of peers on the binary Kademlia tree. L_i peaks close to the theoretical value 2 of an optimally balanced tree. Only few peers leak significantly more. While the mean for R_i is the same, the distribution is slightly different.

Scalability: Load Distribution



(a) Histogram of # of received responses.



(b) Histogram of # of given responses.

In a simulation with n $= 1000$, the number of given (b) and received (a) responses has been recorded for every peer. While the distribution of received responses is very sharp, the distribution for given responses is twice as broad. In the Kademlia routing tables, some peers are more often represented than others.

Dealing with Dishonest Peers

What if peers provide manipulated aggregates?

Assumption

The majority of peers is honest.

- random attribution of peers to leaf nodes
- require signatures on aggregates
- conflicting signatures of P_i constitute proof of deviation:
  - signatures of 2 distinict (inital) aggregates from same peer
  - signatures on parent aggregates that are not computed from child aggregates
- proofs lead to ban of peers and are stored in the DHT
- in case of diverging aggregates:
  take aggregate with most signatures after sampling

Blind Signatures for Authorisation

A	Authority
P_i	Peer, i-th out of n
a_i	Aggregate of P_i
$\sigma_i(m)$	P_i's signature scheme using its key pair (pk_i, sk_i)
$\sigma_A(m)$	Authority's signature scheme
$\chi(m, r)$	Blinding technique with random number r
$\delta(s, r)$	Retrieving technique of blind signature

- P_i provides $b_i = \chi(pk_i, r_i)$ to A
- A provides once for P_i the blinded signature $s_i = \sigma_A(b_i)$
- P_i retrieves authorisation token $t_i = \delta(s_i, r_i)$

Local Validity of Aggregate Signatures

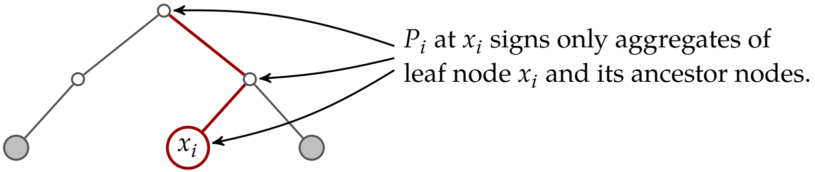


Fig. Eligibility of signatures in ADVOKAT. The public key pk_i of P_i is tied by its authorisation token t_i to one leaf node $x_i = \eta(t_i)$. Signatures of P_i are only valid for aggregates of node x_i and its ancestor nodes.

Eligibility

Proof of Eligiblity

pk_i and its signature t_i from A

Proving Aggregate Authorship of a:

- generate signature for a_i and its properties $p(a)$: $s_a = \sigma_i(\eta(a), p(a))$ with hashing function $\eta(\cdot)$

Proof of Auhorship

a, $p(a)$, s_a, and proof of eligibility pk_i, t_i

Confirmation Requests

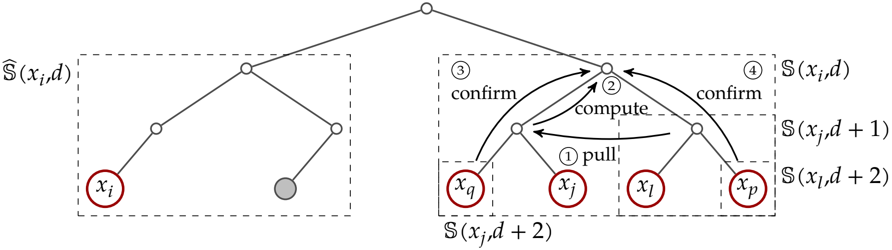


Fig. Pull and confirm of aggregates in ADVOKAT. P_j with x_j produces a confirmed aggregate container of $\mathbb{S}(x_i, d) = \widehat{\mathbb{S}}(x_j, d)$. This scheme applies to all tree levels with possibly large subtrees with multiple potential sources.

- aggregates are confirmed by
  up to 5 signatures from up to 3 peers

Protocol Properties (no formal proofs)

- eligibility of peers
- probabilistic correctness of the root aggregate
- probabilistic confidentiality of initial aggregates
- probabilistic fairness
- verifiability (similar to paper-based voting)
- average number of operations/messages per peer: $\log(n)$

Applications of ADVOKAT

Fig. Online Voting, © Flickr/european_parliament (CC by-nd-nd)

Fig. Online Lottery, Screenshot of `https://www.euro-millions.com`

ADVOKAT-Vote: Protocol

Preparation

- sponsor defines vote (question, peers p_i, authority A) and sends invitations
- each P_i creates (pk_i, sk_i)
- P_i sends authorization request with blinded pk_i to A

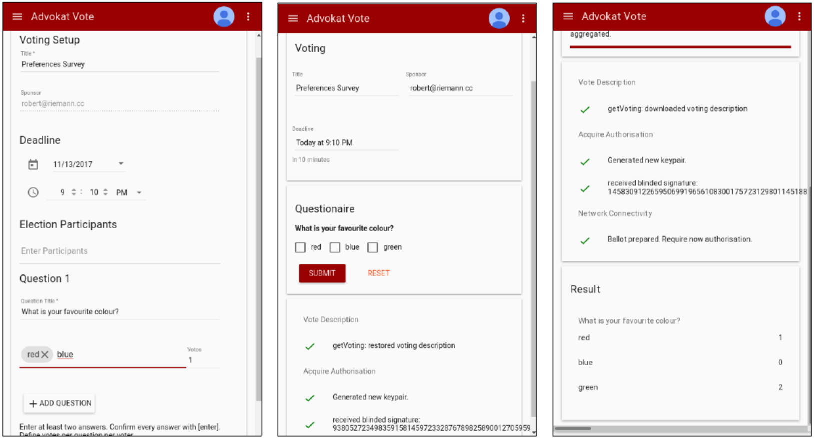
Authorisation

- once for each P_i, A signs blinded $b_i = \chi(pk_i, r_i)$ and sends $s_i = \sigma_A(b_i)$ back
- peers compute authorisation token $t_i = \delta(s_i, r_i)$

Aggregation

- peer P_i joins the Kademlia DHT at $x_i = \eta(t_i)$
- P_i assigns initial aggregate a_i to leaf node x_i
- all peers compute collectively the root aggregate a_R using ADVOKAT

ADVOKAT-Vote: Implementation



(a) Setup Phase (b) Preparation Phase (c) Evaluation Phase

Fig. Demonstrator implemented in HTML/JS using WebRTC

Online Lottery: Challenge

Neither players nor the authority shall estimate the outcome as long as tickets are sold.

ADVOKAT-Lottery: Online Lottery[12]

Ticket Purchase

- each P_i generates (pk_i, sk_i) and picks number r_i
- P_i buys authorisation from A and receives t_i
- P_i joins Kademlia DHT with $x_i = \eta(t_i)$

Distributed Random Process (Aggregation)

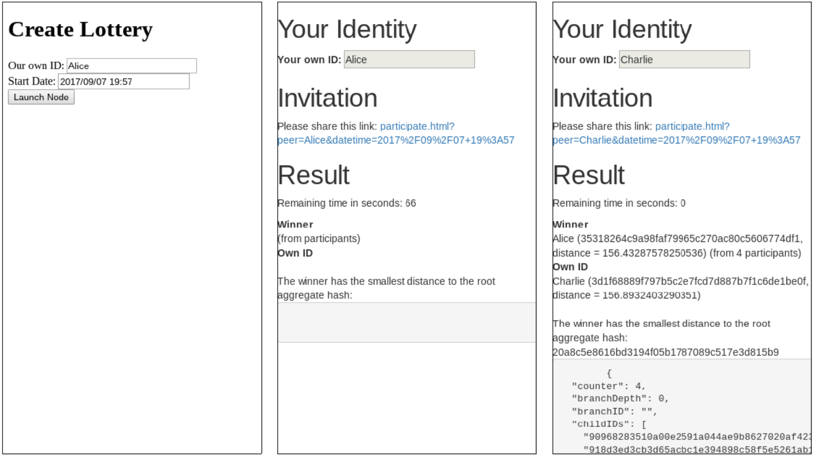
- peers compute jointly the Merkle root a_R of all $a_i = \mathrm{commitment}(r_i)$

Winner Identification

- A learns a_R by sampling
- Winners from list ordered by x_i `XOR` a_R

[12] R. Riemann and S. Grumbach. "Distributed Random Process for a large-scale Peer-to-Peer Lottery". In: Proc. of 17th IFIP Distributed Applications and Interoperable Systems. DAIS'17. Neuchâtel: Springer, June 2017, pp. 34–48. DOI: 10.1007/978-3-319-59665-5_3.

ADVOKAT-Lottery: Implementation



(a) Setup Phase (b) Preparation Phase (c) Evaluation Phase

Fig. Demonstrator implemented in HTML/JS using WebRTC

Conclusion

Distributed protocols are promising for trustworthy aggregation protocols.

- proposed new protocol ADVOKAT
- new compromise to balance: verifiability and confidentiality
- new approach to trust in technology: bring your own, reduced complexity
- new privacy-enhancing tool (PET) for privacy by design
- various potential use-cases: voting, lottery, health data, auctions, sensor data, etc.

Thesis Statement

We claim that distributed protocols are promising to carry out trustworthy aggregations of confidential data.

French National Assembly Flickr/partisocialiste (CC by-nc-nd)

Voting Protocols

Fig. Online Voting © Flickr/european_parliament (CC by-nd-nd)

Fig. Paper-based Voting © Flickr/coventrycc (CC by-nd-nd)

Complexity of Cooperation

Observations

1. size of cooperation is increasing in terms of peers & links
2. diversification and specialisation
3. overall complexity is increasing

Problems

1. How to ensure trust in cooperation?
2. How to govern large cooperations?

Generic Paper-based Voting

1. **Preparation Phase**
   central voter registry issues list of eligible voters,
   prints undistinguishable voting ballots
2. **Casting Phase**
   on-site, public supervision, voting station(s) run by citizens
3. **Aggregation Phase**
   tallying of casted ballots
4. **Evaluation Phase**
   computation of the voting outcome from public tally
5. **Verification Phase**
   observation during the vote (eye-sight), recounts

Challenge: Conflicting Protocol Properties

Ensure set of security properties at the same time:

- unconditional secrecy of the ballot
- universal verifiability of the tally
- eligibility of the voter

Achievable only with unrealistic assumptions[13]:

compromise required

[13] B. Chevallier-Mames et al. "On Some Incompatible Properties of Voting Schemes". In: Towards Trustworthy Elections: New Directions in Electronic Voting. Springer, 2010.

Technology Impact on Voting I

Fig. Digital Natives.
© Flickr/antmcneill (CC by-sa)

Fig. Paper-based Voting.
© Flickr/coventrycc (CC by-nd-nd)

Technology Impact on Voting II

Impact on Expectations

- comfort on a par with other online services
- flexibility
- automation for cost efficiency

Impact on Security

- hidden body cameras
- invisible ink
- fingerprint databases
- DNA analysis

Online Voting

Online Voting

remote electronic voting

- no chain of custody verifiable per eye-sight
- electronic signals are easy to duplicate

Need for new concepts to ensure security properties.

Empowerment of Voters

Assumption of a Distributed Online Voting Protocol

- no authority
- equally privileged, equipotent voters

Promises

- reflects democratic principle of equally powerful voters
- all voters are potential voting officers
- all voters responsible to enfore policy of protocol
- with no weakest link, promise of improved resiliance against DDoS attacks
- balance of knowledge among voters

Notions of Distribution in Online Voting

1. **Degree of Specialisation**
   from equipotent voters to specialised authorities
2. **Topology** of communication/responsabilities
   from centralised over decentralised to distributed
3. **Phase**
   consider phases that are actually distributed

Notions of Distribution in Online Voting

1. **Degree of Specialisation**
   from equipotent voters to specialised authorities
2. **Topology** of communication/responsabilities
   from centralised over decentralised to distributed
3. **Phase**
   consider phases that are actually distributed

Fully distributed Protocol

- equipotent voters, no authorities,
- distributed topology
- in all phases (but the registration)

Online Lottery

Requirements on Online Lottery:

- correctness of random process
- verifiability of random process
- privacy of the (winning) player
- validity of the ticket (eligibility)
- confidentiality of the ticket number
- completeness of the reward

Paper-based Lottery

- players buy tickets from Authority in person
- player verify random nature of drawing setup
- winning tickets are drawn from urn under public supervision of all players
- all other tickets are drawn to convince the loosers of the correctness
- random process cannot be repeated

ADVOKAT as Middleware

Distributed Aggregation of Confidential Data:

- **Online Voting**
- **Online Lottery**
- Auctions
- Personal Data, especially Health Data
- Sensor Data

Blind RSA Signatures

- $m' = mr^e \mod N$
- $s' = (m')^d \mod N$
- $s = s' \cdot r^{-1} \mod N = m^d \mod N$
  with $r^{ed} = r \mod N$