Large-scale trustworthy distributed collaborative systems

Claudia-Lavinia Ignat
Coast team
Inria Nancy-Grand Est

Habilitation defense
April 23, 2021

claudia.ignat@inria.fr

Education and work experience

Technical University
Cluj-Napoca Romania

- Master thesis ETH Zurich, June 2000
- Engineer diploma, July 2000

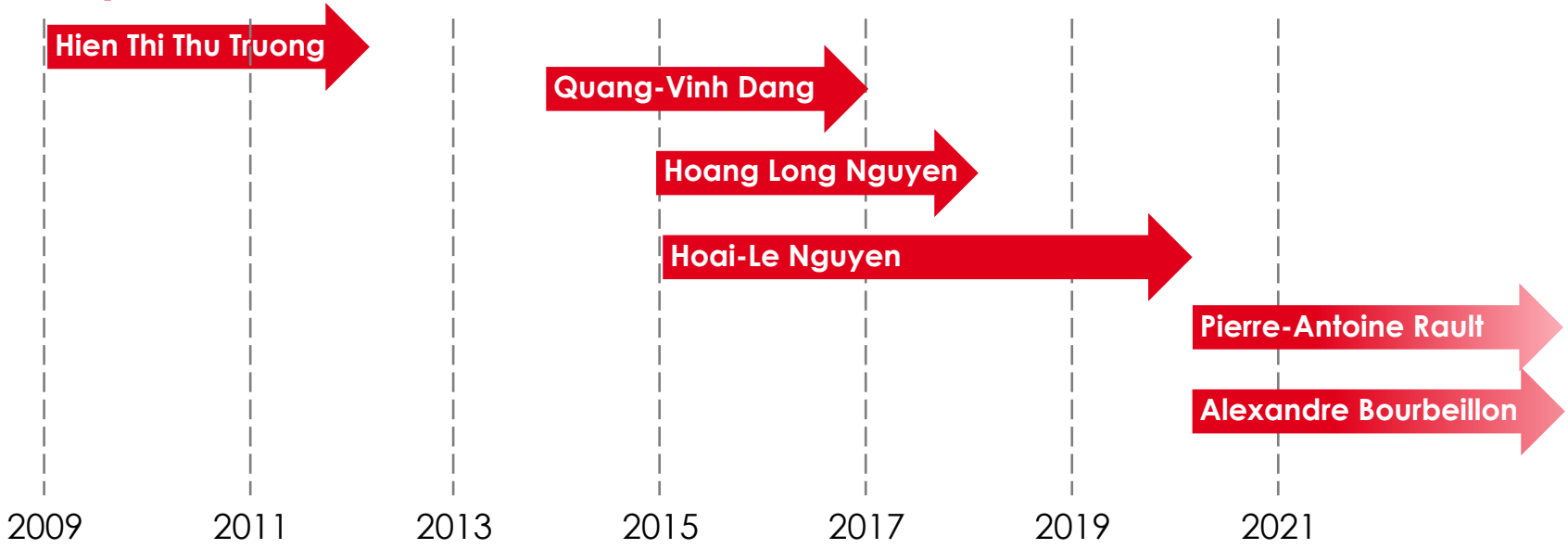
ETH Zurich
Switzerland

- Research and teaching assistant
- PhD thesis July 2006

Inria Nancy-Grand Est
France

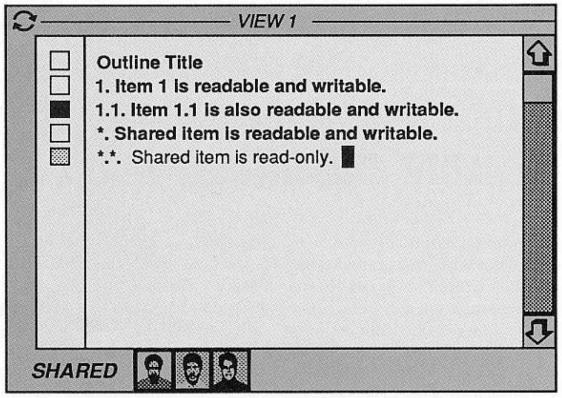
- Postdoc 2006-2007
- Research scientist since 2007

PhD Supervision



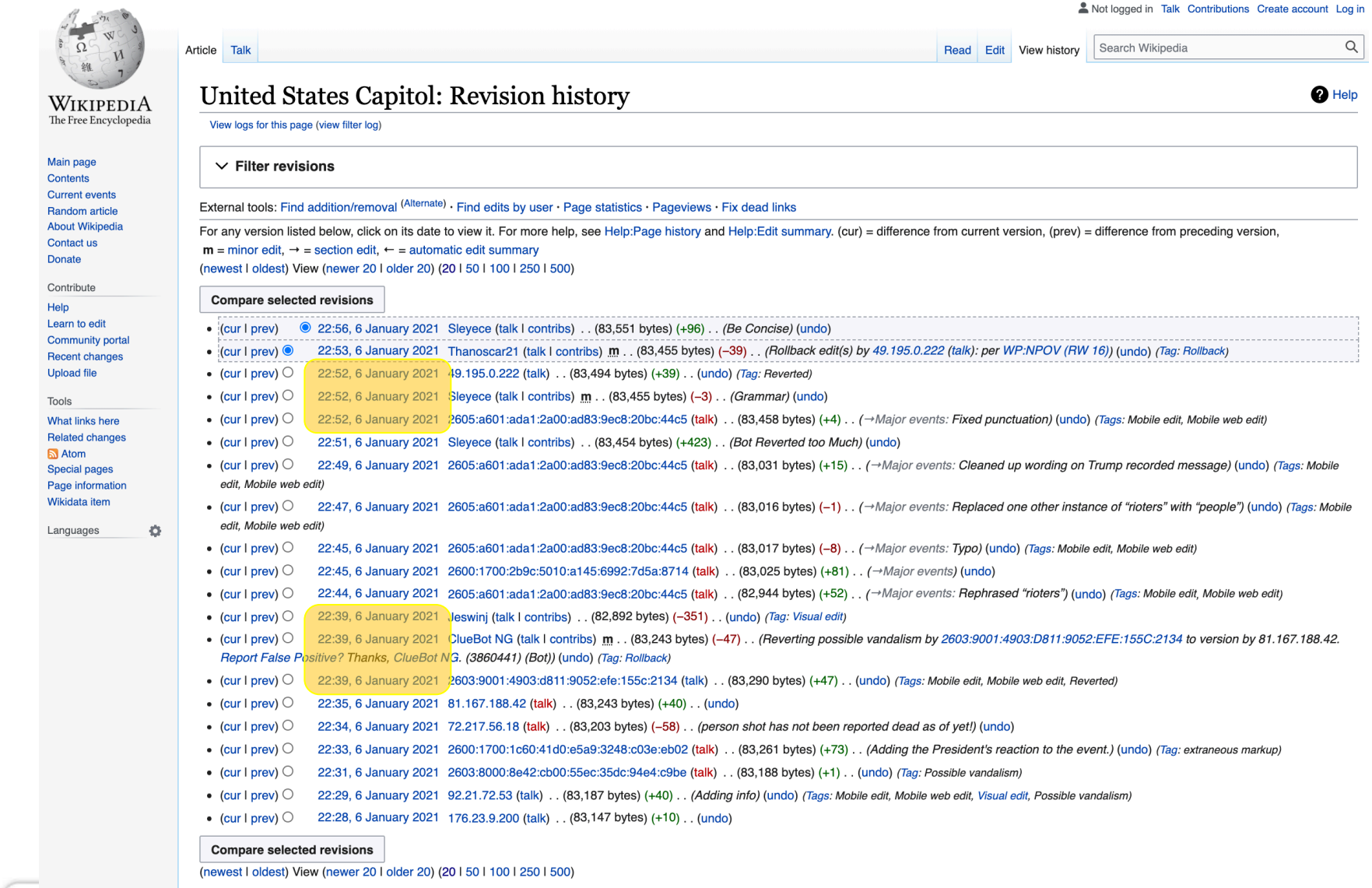
Collaborative Systems: from users to community of users

GROVE, 1989



"Why would a group ever want to edit in the same line of text at the same time?" [EGR91]

From users to community of users: new practices



Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools
What links here
Related changes
Atom
Special pages
Page information
Wikidata item

Languages

United States Capitol: Revision history

Help

View logs for this page (view filter log)

Filter revisions

External tools: Find addition/removal (Alternate) · Find edits by user · Page statistics · Pageviews · Fix dead links

For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary. (cur) = difference from current version, (prev) = difference from preceding version, **m** = minor edit, → = section edit, ← = automatic edit summary

(newest | oldest) View (newer 20 | older 20) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- (cur | prev) 22:56, 6 January 2021 Sleyece (talk | contribs) . . (83,551 bytes) (+96) . . *(Be Concise)* (undo)
- (cur | prev) 22:53, 6 January 2021 Thanoscar21 (talk | contribs) **m** . . (83,455 bytes) (−39) . . *(Rollback edit(s) by 49.195.0.222 (talk): per WP:NPOV (RW 16))* (undo) *(Tag: Rollback)*
- (cur | prev) 22:52, 6 January 2021 49.195.0.222 (talk) . . (83,494 bytes) (+39) . . (undo) *(Tag: Reverted)*
- (cur | prev) 22:52, 6 January 2021 Sleyece (talk | contribs) **m** . . (83,455 bytes) (−3) . . *(Grammar)* (undo)
- (cur | prev) 22:52, 6 January 2021 2605:a601:ada1:2a00:ad83:9ec8:20bc:44c5 (talk) . . (83,458 bytes) (+4) . . *(→Major events: Fixed punctuation)* (undo) *(Tags: Mobile edit, Mobile web edit)*
- (cur | prev) 22:51, 6 January 2021 Sleyece (talk | contribs) . . (83,454 bytes) (+423) . . *(Bot Reverted too Much)* (undo)
- (cur | prev) 22:49, 6 January 2021 2605:a601:ada1:2a00:ad83:9ec8:20bc:44c5 (talk) . . (83,031 bytes) (+15) . . *(→Major events: Cleaned up wording on Trump recorded message)* (undo) *(Tags: Mobile edit, Mobile web edit)*
- (cur | prev) 22:47, 6 January 2021 2605:a601:ada1:2a00:ad83:9ec8:20bc:44c5 (talk) . . (83,016 bytes) (−1) . . *(→Major events: Replaced one other instance of "rioters" with "people")* (undo) *(Tags: Mobile edit, Mobile web edit)*
- (cur | prev) 22:45, 6 January 2021 2605:a601:ada1:2a00:ad83:9ec8:20bc:44c5 (talk) . . (83,017 bytes) (−8) . . *(→Major events: Typo)* (undo) *(Tags: Mobile edit, Mobile web edit)*
- (cur | prev) 22:45, 6 January 2021 2600:1700:2b9c:5010:a145:6992:7d5a:8714 (talk) . . (83,025 bytes) (+81) . . *(→Major events)* (undo)
- (cur | prev) 22:44, 6 January 2021 2605:a601:ada1:2a00:ad83:9ec8:20bc:44c5 (talk) . . (82,944 bytes) (+52) . . *(→Major events: Rephrased "rioters")* (undo) *(Tags: Mobile edit, Mobile web edit)*
- (cur | prev) 22:39, 6 January 2021 Jeswinj (talk | contribs) . . (82,892 bytes) (−351) . . (undo) *(Tag: Visual edit)*
- (cur | prev) 22:39, 6 January 2021 ClueBot NG (talk | contribs) **m** . . (83,243 bytes) (−47) . . *(Reverting possible vandalism by 2603:9001:4903:D811:9052:EFE:155C:2134 to version by 81.167.188.42. Report False Positive? Thanks, ClueBot NG. (3860441) (Bot))* (undo) *(Tag: Rollback)*
- (cur | prev) 22:39, 6 January 2021 2603:9001:4903:d811:9052:efe:155c:2134 (talk) . . (83,290 bytes) (+47) . . (undo) *(Tags: Mobile edit, Mobile web edit, Reverted)*
- (cur | prev) 22:35, 6 January 2021 81.167.188.42 (talk) . . (83,243 bytes) (+40) . . (undo)
- (cur | prev) 22:34, 6 January 2021 72.217.56.18 (talk) . . (83,203 bytes) (−58) . . *(person shot has not been reported dead as of yet!)* (undo)
- (cur | prev) 22:33, 6 January 2021 2600:1700:1c60:41d0:e5a9:3248:c03e:eb02 (talk) . . (83,261 bytes) (+73) . . *(Adding the President's reaction to the event.)* (undo) *(Tag: extraneous markup)*
- (cur | prev) 22:31, 6 January 2021 2603:8000:8e42:cb00:55ec:35dc:94e4:c9be (talk) . . (83,188 bytes) (+1) . . (undo) *(Tag: Possible vandalism)*
- (cur | prev) 22:29, 6 January 2021 92.21.72.53 (talk) . . (83,187 bytes) (+40) . . *(Adding info)* (undo) *(Tags: Mobile edit, Mobile web edit, Visual edit, Possible vandalism)*
- (cur | prev) 22:28, 6 January 2021 176.23.9.200 (talk) . . (83,147 bytes) (+10) . . (undo)

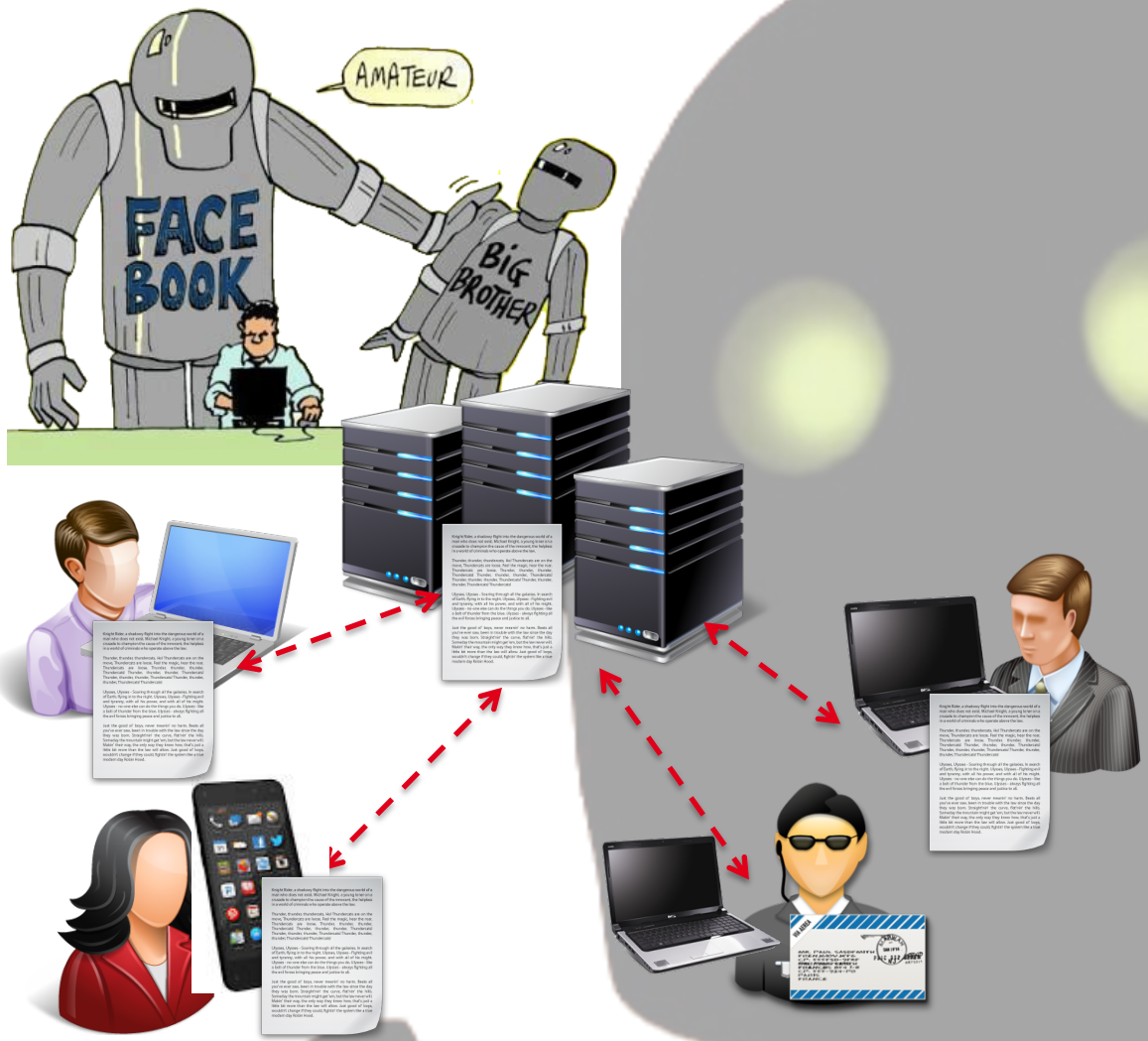
Compare selected revisions

(newest | oldest) View (newer 20 | older 20) (20 | 50 | 100 | 250 | 500)

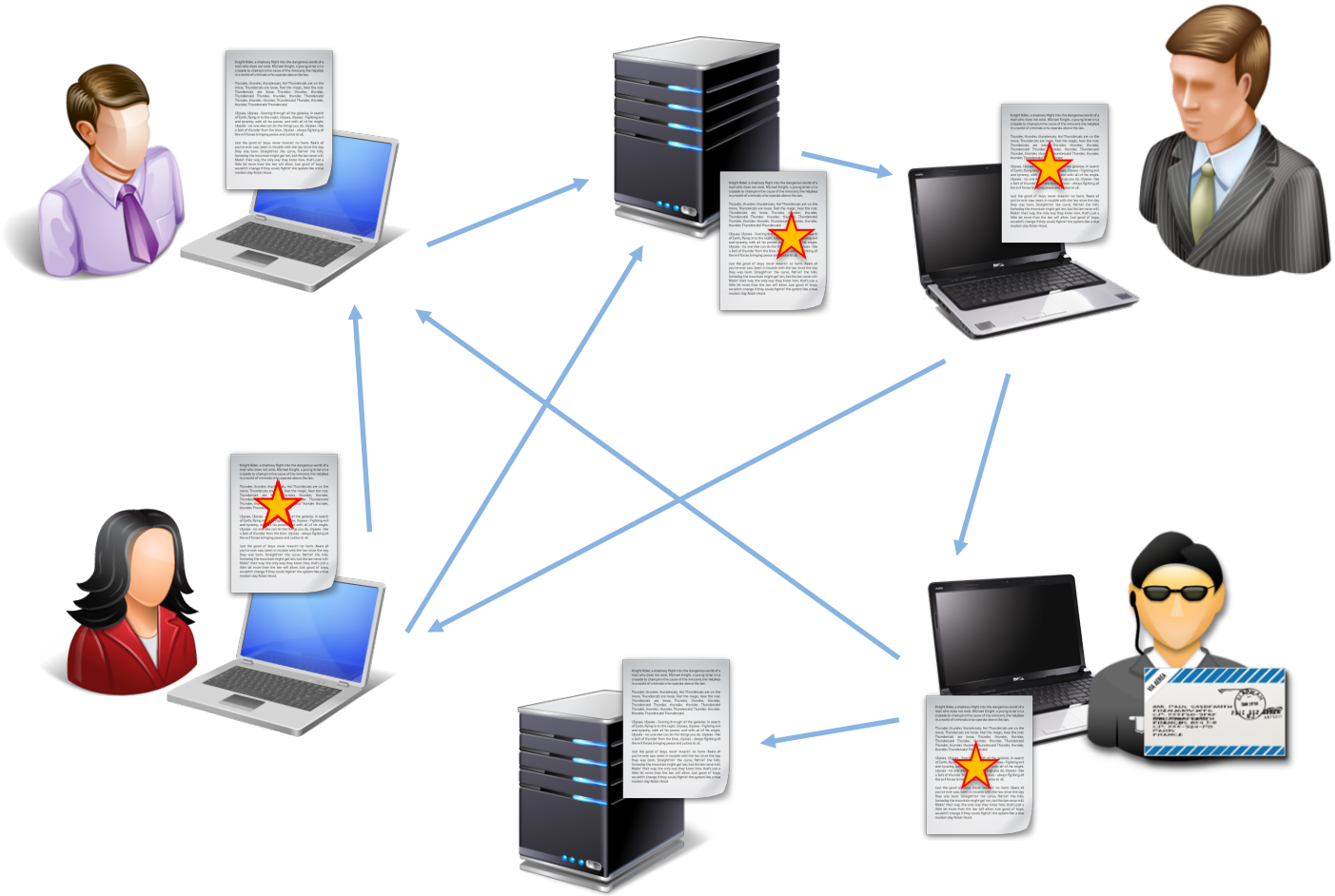
Limitations of Central Authority Systems

SCALABILITY

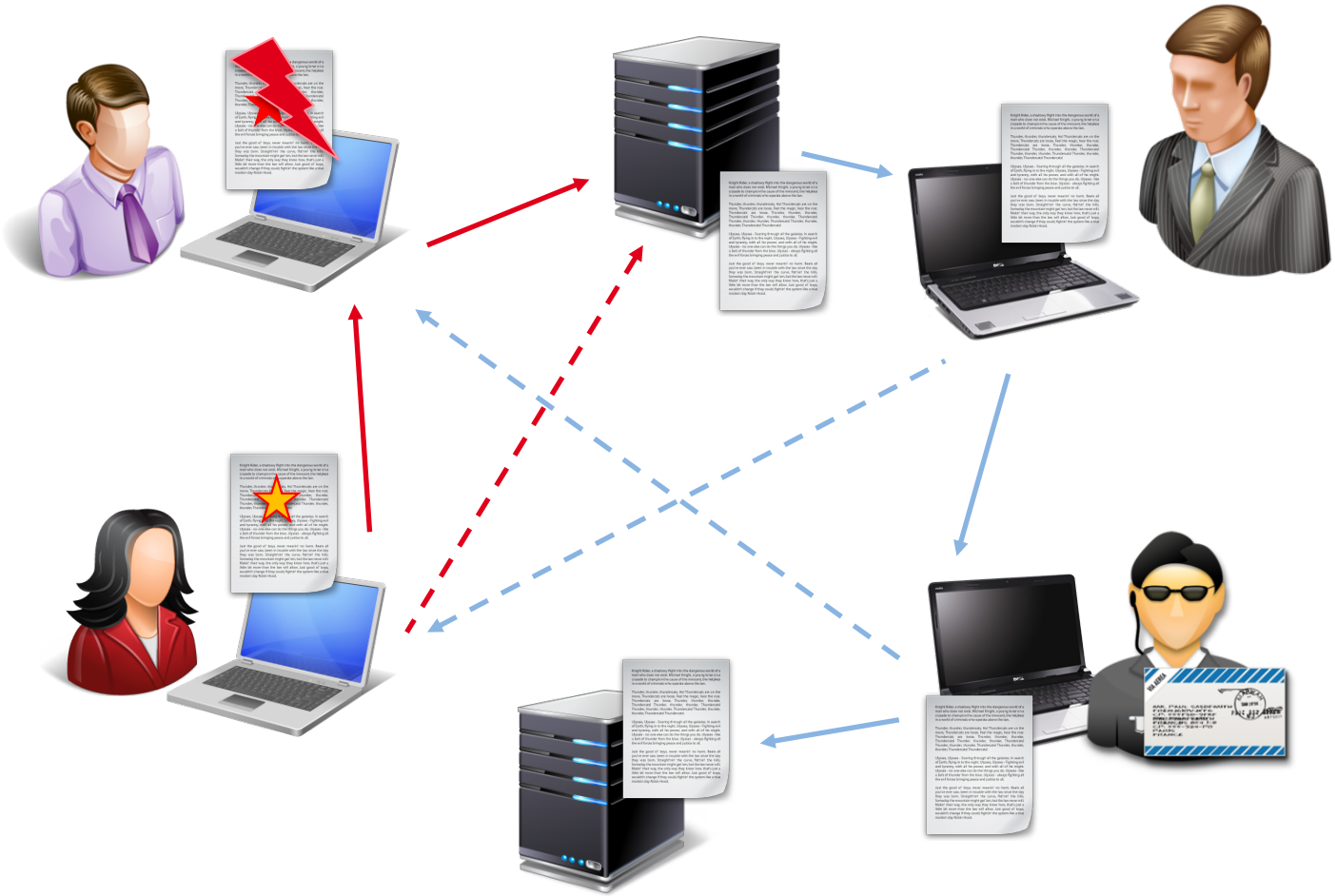
PRIVACY



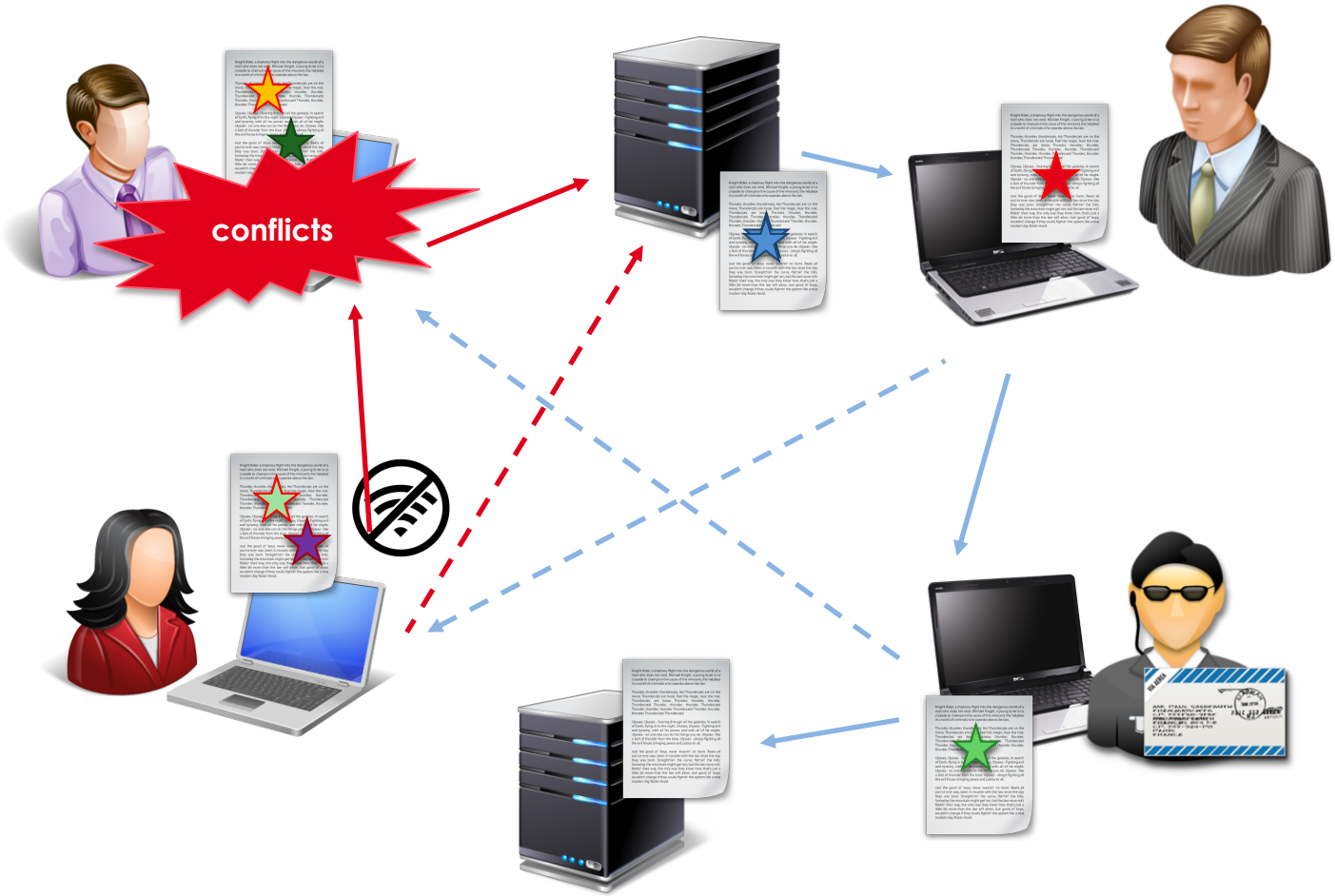
Peer-to-Peer Collaborative Systems



Collaboration Modes – Concurrent Changes

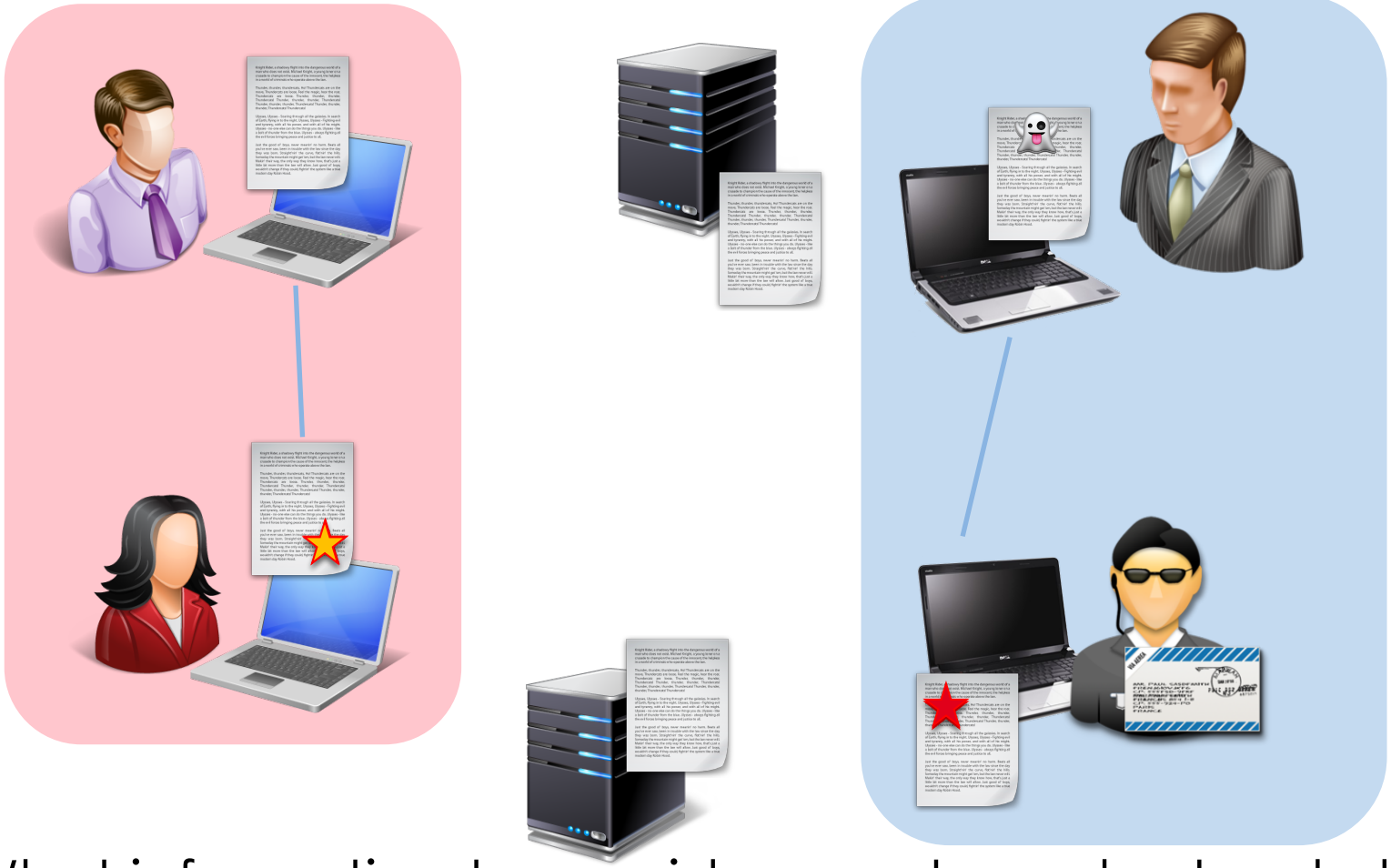


Collaboration Modes – Offline Work



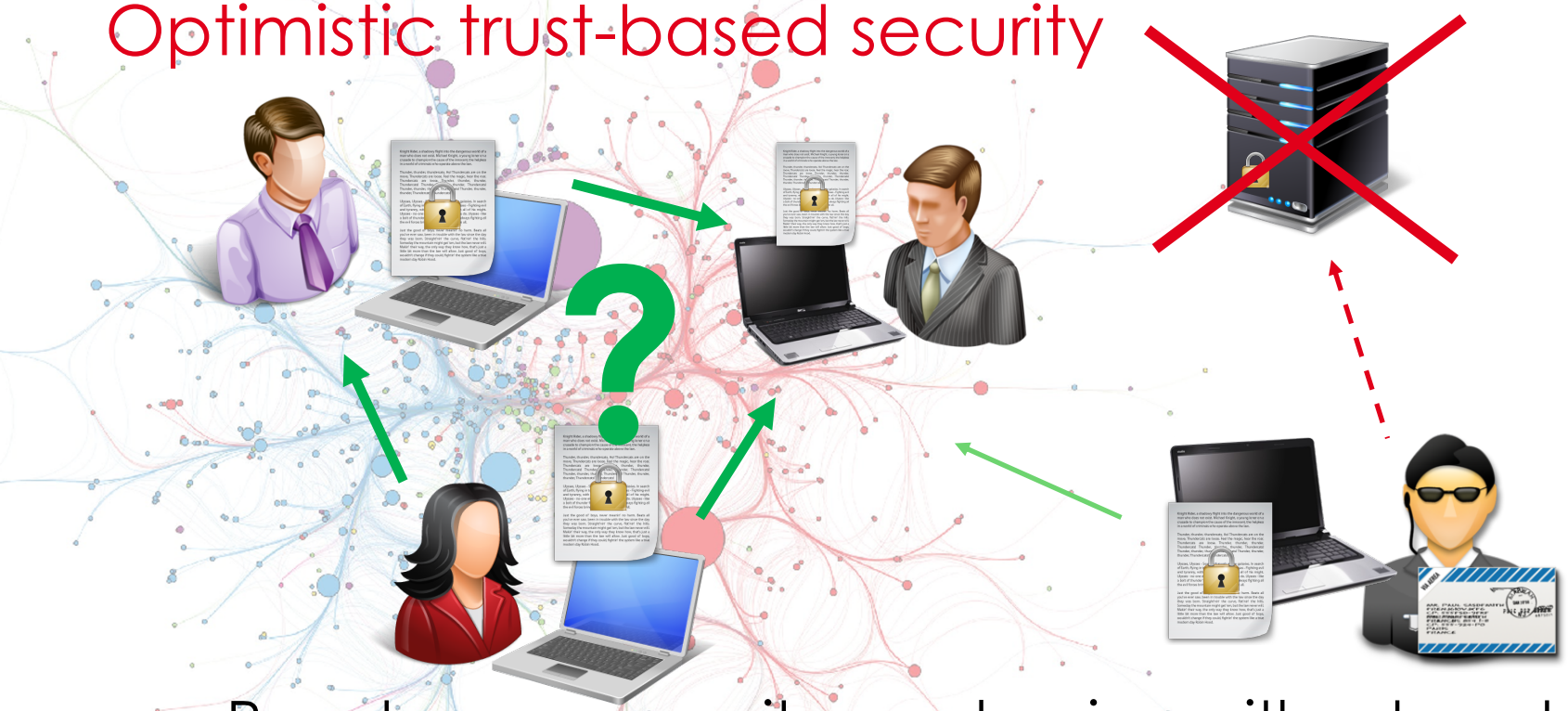
Collaboration Modes – Ad-hoc Collaboration

Group Awareness



What information to provide users to understand changes of other users and prevent conflictual changes?

Optimistic trust-based security



- Peer-to-peer security mechanism without central authority, scalable and easy-to-use
- Soft security: give access to data without control but with a-posteriori verified restrictions
- Trust-based data access with dynamic trust based on collaborative experience

Research Axes

1. Collaborative Data Management
   - Reliable, scalable and explainable replication algorithms
   - Evaluation: collaboration traces and user studies
   - Group awareness mechanisms

2. Trustworthy Collaboration
   - Evaluate collaborators trust based on their past behavior
   - Evaluation: game theory and user studies

Research Axes

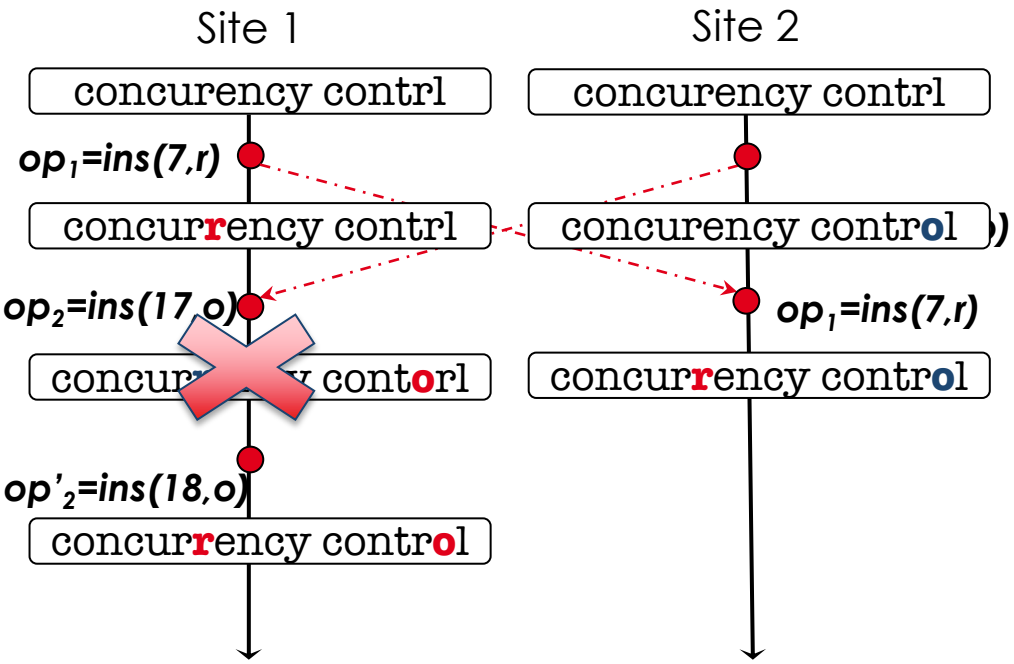
1 Collaborative Data Management

- Reliable, scalable and explainable replication algorithms
- Evaluation: collaboration traces and user studies
- Group awareness mechanisms

2 Trustworthy Collaboration

- Evaluate collaborators trust based on their past behavior
- Evaluation: game theory and user studies

Operational transformation (OT) [EG89]



T(**ins(p1,c1)**, **ins(p2,c2)**) :-
if (**p1<p2**) return **ins(p1,c1)**
else return **ins(p1+1,c1)**
endif

- Transforms non commuting operations to make them commute
- Genericity
- Time complexity
  Average: O(H.c) H: #ops
  Worst case: $O(H^2)$ c: avg. #conc. ops
- Difficult to write correct transformation functions
- State vectors used for detecting concurrency ⇒ scalability limitations
- **Not very suitable for large scale peer-to-peer collaboration**

Conflict-free Replicated Data Types (CRDT) [SPBZ11]

- Design operations to be commutative by construction
- Document = linear sequence of elements
  - Each element has a unique identifier
  - Identifier constant for the lifetime of the document
  - Dense total order of identifiers consistent with element order:
    - $\forall\, id_x , id_y : id_x < id_y \Rightarrow \exists\, id_z : id_x < id_z < id_y$
- Different approaches for generating identifiers

Conflict-free Replicated Data Types (CRDT)
Logoot [WUM09]

- Logoot identifiers: $\langle p_1,s_1,h_1\rangle\langle p_2,s_2,h_2\rangle \cdots \langle p_k,s_k,h_k\rangle$

  p_i position

  s_i site identifier

  h_i logical clock at site s_i

<1,2,1>	c
<1,2,2>	o
<2,1,2>	n
<3,1,3>	c
<3,1,3><8,4,5>	u
<3,2,5>	r
<4,1,7>	e
<4,1,7><9,2,6>	n
<7,2,8>	c
<9,1,7>	y
<10,2,8>	
<12,3,1>	c
<12,3,1><6,5,1>	o
<12,3,1><7,8,2>	n
<12,3,1><7,8,2><12,3,5>	t
<12,3,1><7,8,2><13,3,6>	r
<12,3,1><7,8,2><14,3,7>	l

ins(<3,2,5><13,1,7>, r)

ins(<12,3,1><7,8,2><13,3,6><7,2,9>, o)

- Time complexity
  - Average: O(k log(n))
  - Worst case: O(H*log(H))

  H: #ops
  n: doc. size (non deleted chars.)
  k: avg. size of Logoot identifier
- No need for concurrency detection
- Identifiers storage cost
- New design for each data type
- **Suitable for large-scale collaboration**

My Contributions

- Algorithms design
  - OT algorithms for complex data such as hierarchical text documents [IN08], XML [IO07] and wikis [IA017]
  - A CRDT approach for strings that limits metadata [AMOI13]
  - An undo mechanism based on CRDT [YAI15,YEI19]

- Algorithms evaluation
  - Evaluation of OT and CRDT algorithms [IOMC07,AIOR11]
  - Measurement of delays in real-time collaborative editing systems [DI16a] and their influence on users [IOFS15]

- Conflicts prevention and resolution
  - Awareness mechanisms for source code [IO08] and textual documents that considers user privacy [IPON08]
  - Conflict analysis in git-based projects [NI18] and real-time collaborative editing [NI20]

My Contributions

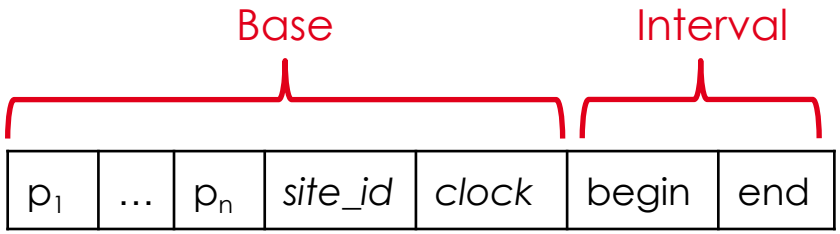
- Algorithms design
  - OT algorithms for complex data such as hierarchical text documents **[IN08]**, XML **[IO07]** and wikis **[IA017]**
  - A CRDT approach for strings that limits metadata **[AMOI13]**
  - An undo mechanism based on CRDT **[YAI15,YEI19]**

- Algorithms evaluation
  - Evaluation of OT and CRDT algorithms **[IOMC07,AIOR11]**
  - Measurement of delays in real-time collaborative editing systems [DI16a] and their influence on users **[IOFS15]**

- Conflicts prevention and resolution
  - Awareness mechanisms for source code **[IO08]** and textual documents that considers user privacy **[IPON08]**
  - Conflict analysis in git-based projects **[NI18]** and real-time collaborative editing **[NI20]**

Conflict-free Replicated Data Types (CRDT)
LogootSplit [AMOI13]

LogootSplit identifiers



1,1,[0,16]	concurency contrl

Insert r between "concur" and "ency contrl"

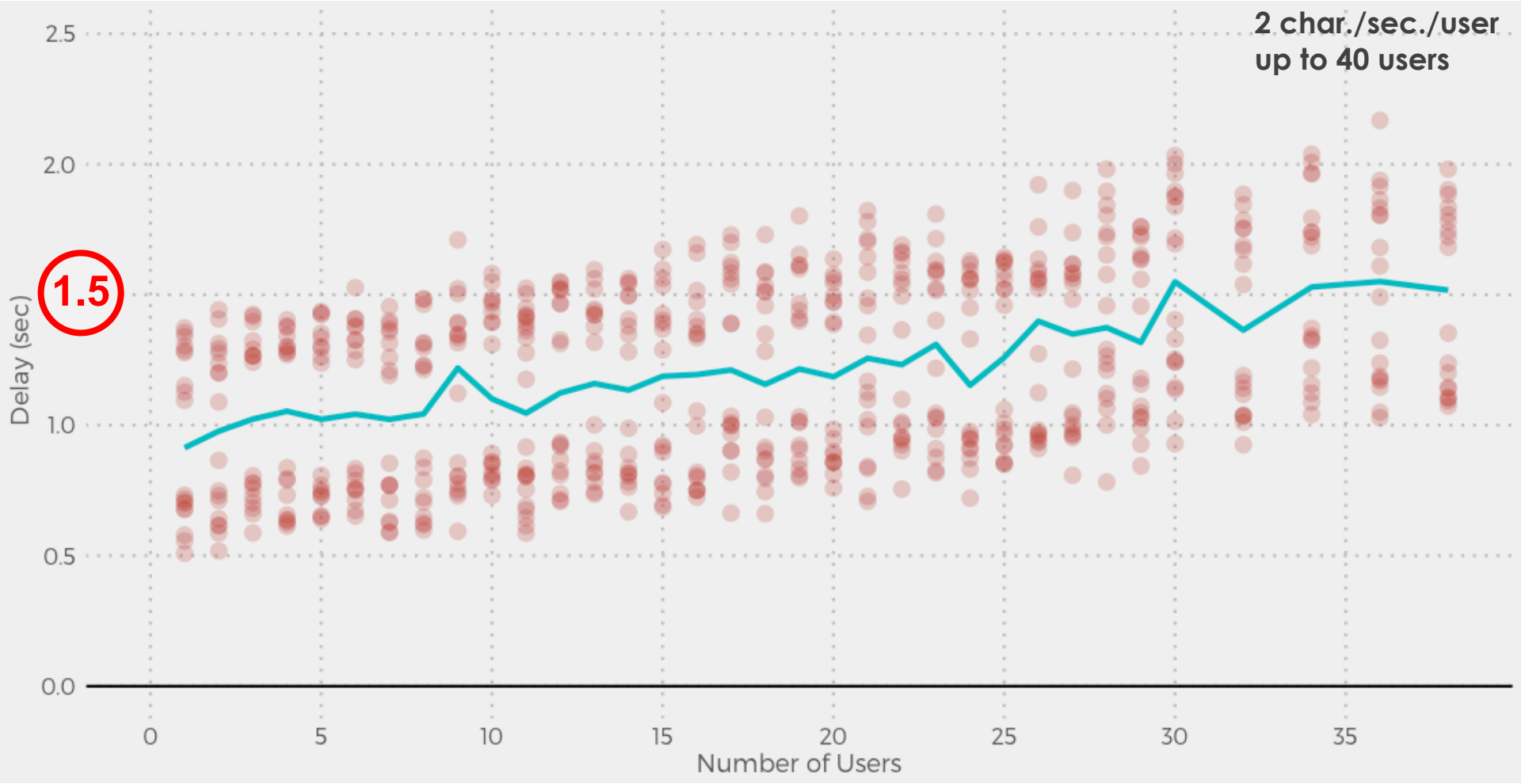
1,1,[0,5]	concur
1,1,5,2,1,[0,0]	**r**
1,1, [6,16]	ency contrl

Insert o between "ency contr" and "l"

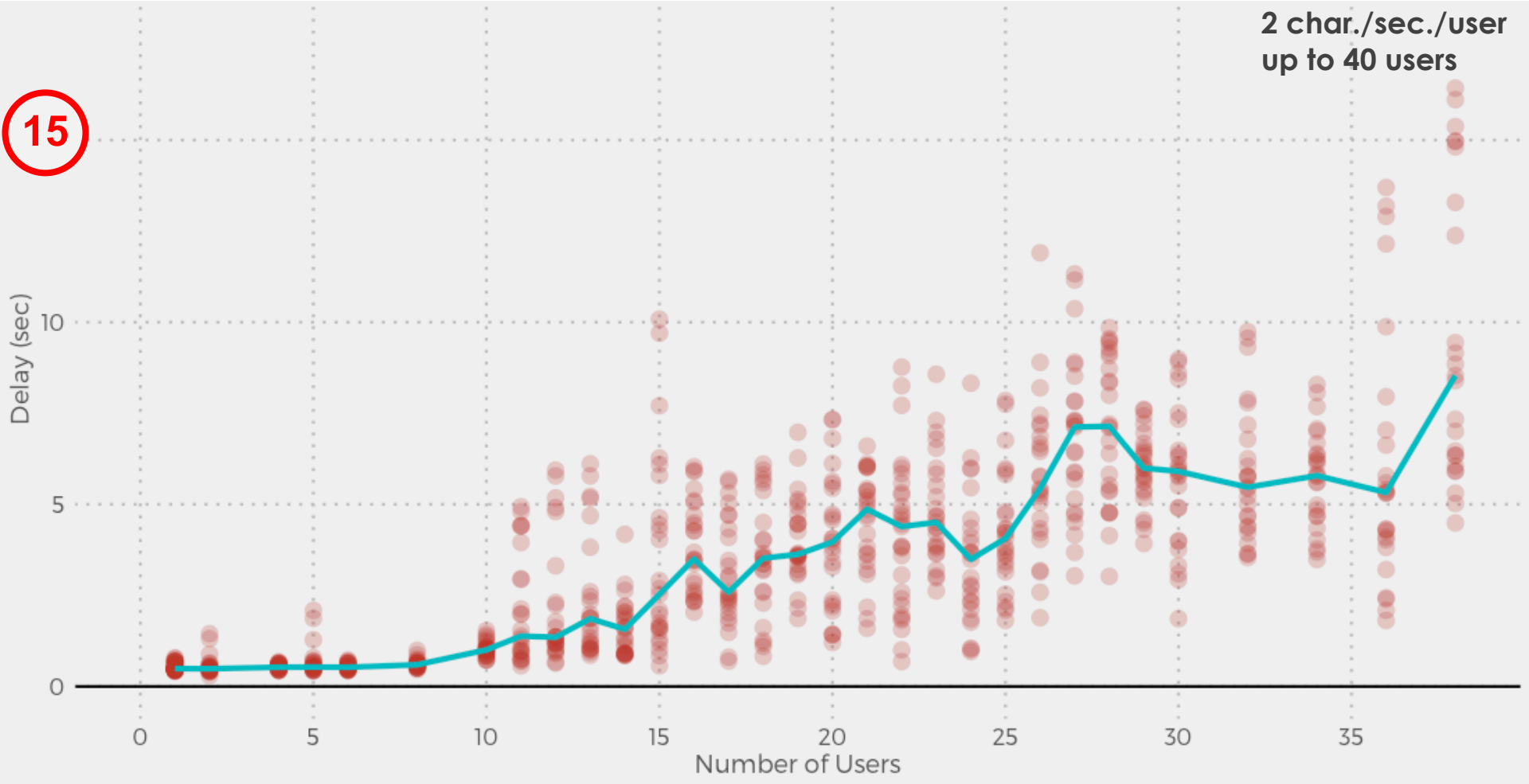
1,1,[0,5]	concur
1,1,5,2,1,[0,0]	**r**
1,1, [6,15]	ency contr
1,1,15,3,1,[0,0]	**o**
1,1,[16,16]	l

Delays in MUTE [NEOIC17]

https://coedit.re/



Delays in GoogleDocs [DI16a]

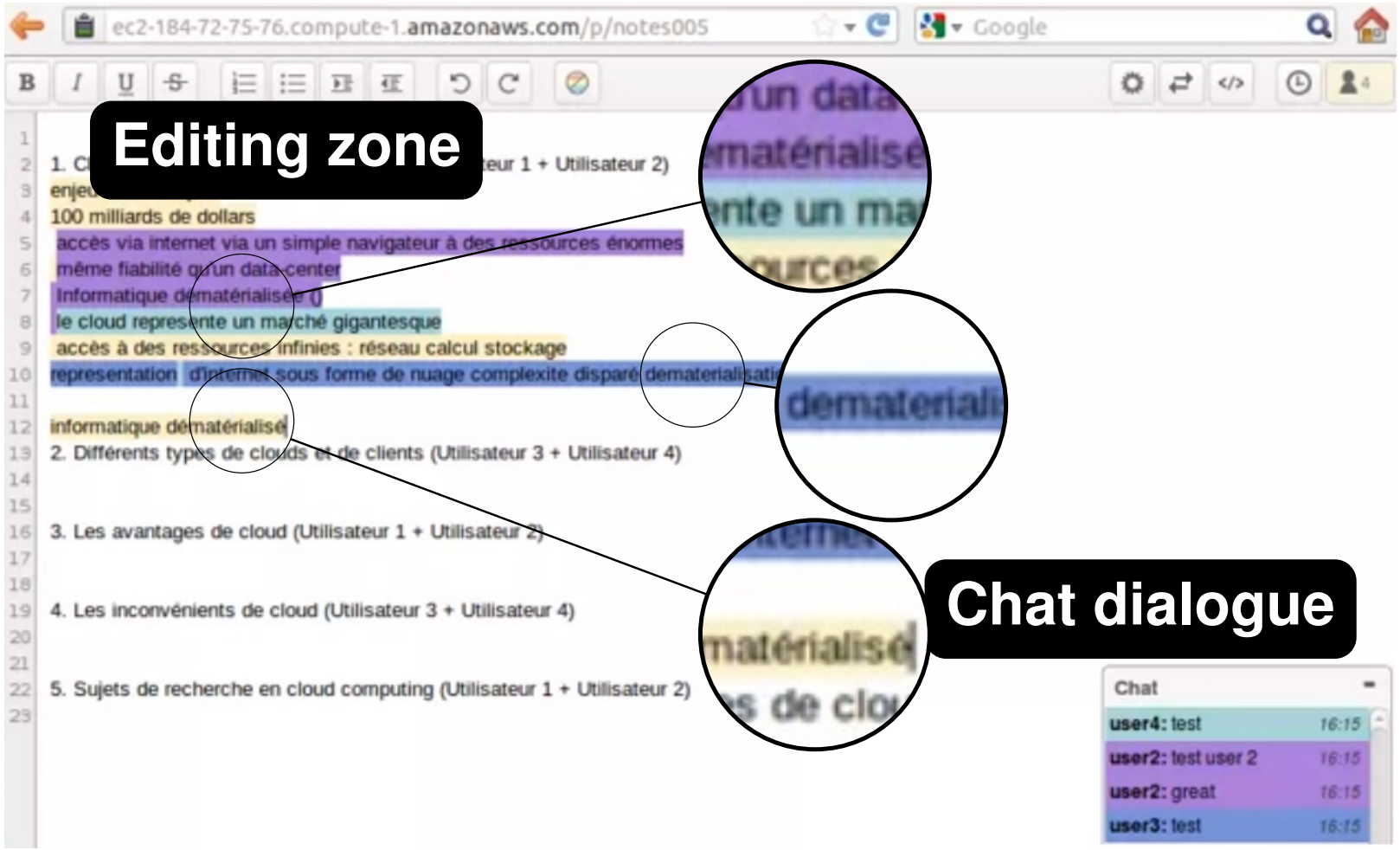


Experimental design: The effect of delay on users

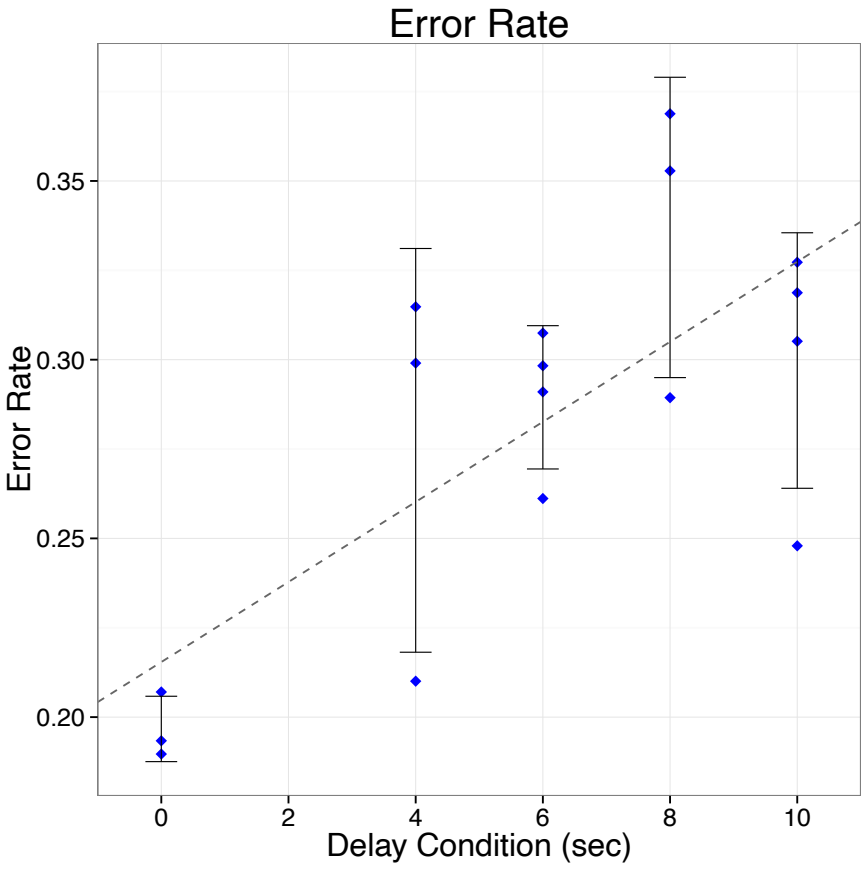
- 20 groups of 4 students
  - Perform several collaborative editing tasks
    - A proofreading task
    - A sorting task
    - **A note taking task**
  - Use the provided collaborative editor (Etherpad) + chat
  - Each group experienced a **certain delay** (0, 4, 6, 8, 10 s)

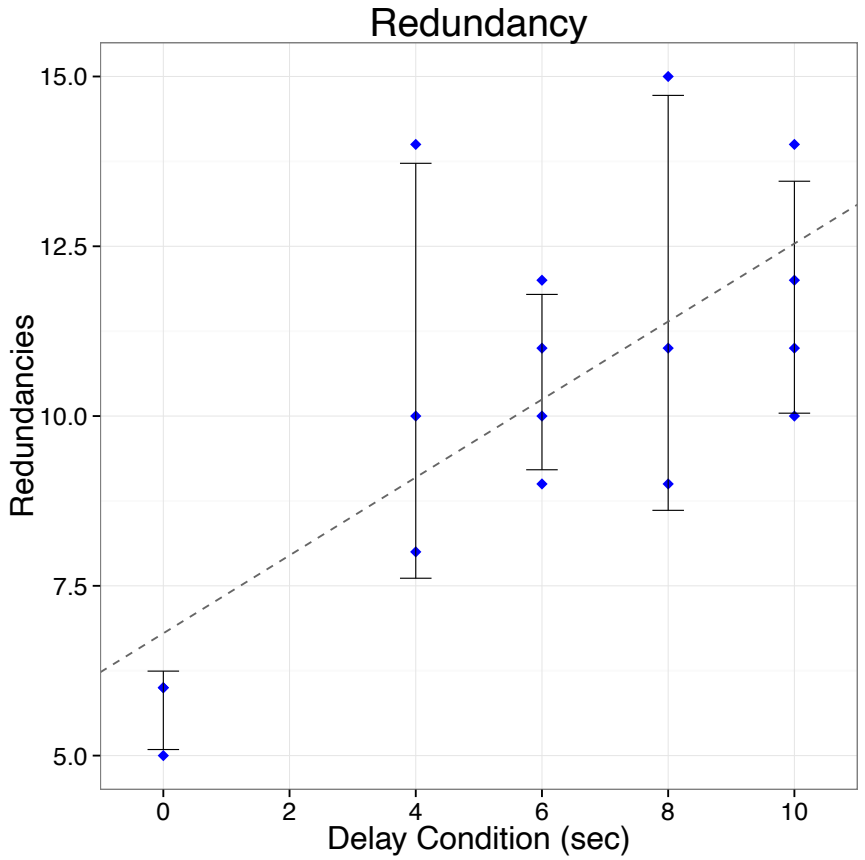
- Collaboration with Department of Psychology Wright State University, Inria USCoast associate team

Note-taking [IOFSC15]



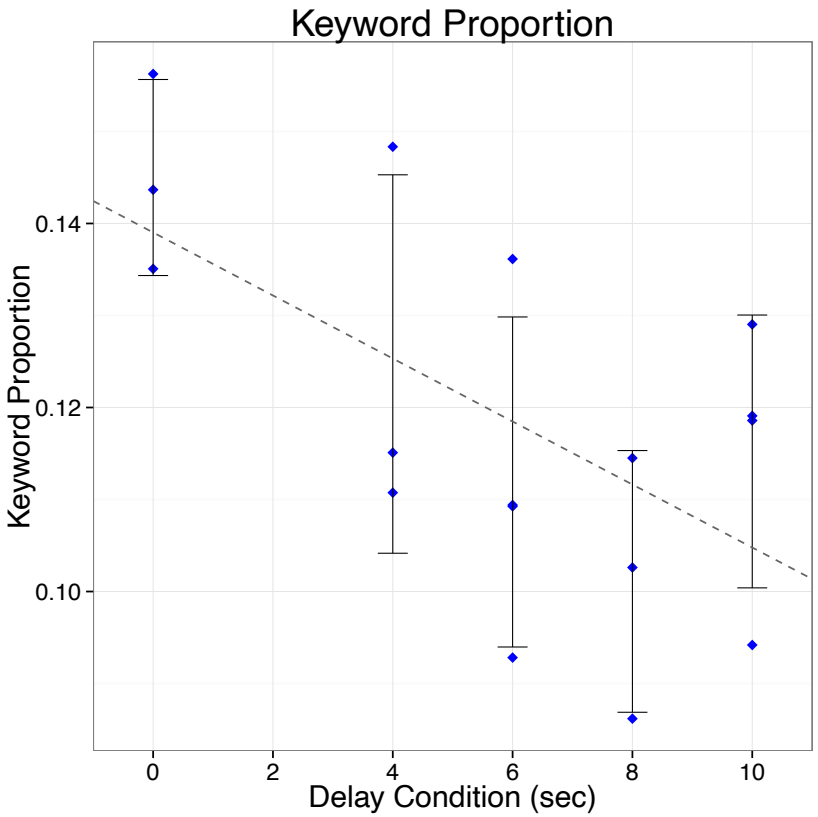
Delay reduces Group Performance





- Delay increases error rate and redundancy

Delay reduces Group Performance



- Delay decreases proportion of keywords

Design implications

- Reduce the delay by the choice of the architecture and synchronisation algorithms
- Make users aware of existing delays such that they can compensate for the delay by coordination strategies

Collaborative Data Management - summary

- Contributions ranging from OT and CRDT **algorithms design** to their **evaluation** (theoretical/simulations/real collaboration traces) and study of their **impacts on users**

- Contributions ranging from the **study of conflicts** in real collaboration traces to **design** and **prototype** of **awareness** mechanisms

Research Axes

1. Collaborative Data Management
   - Reliable, scalable and explainable replication algorithms
   - Evaluation: collaboration traces and user studies
   - Group awareness mechanisms

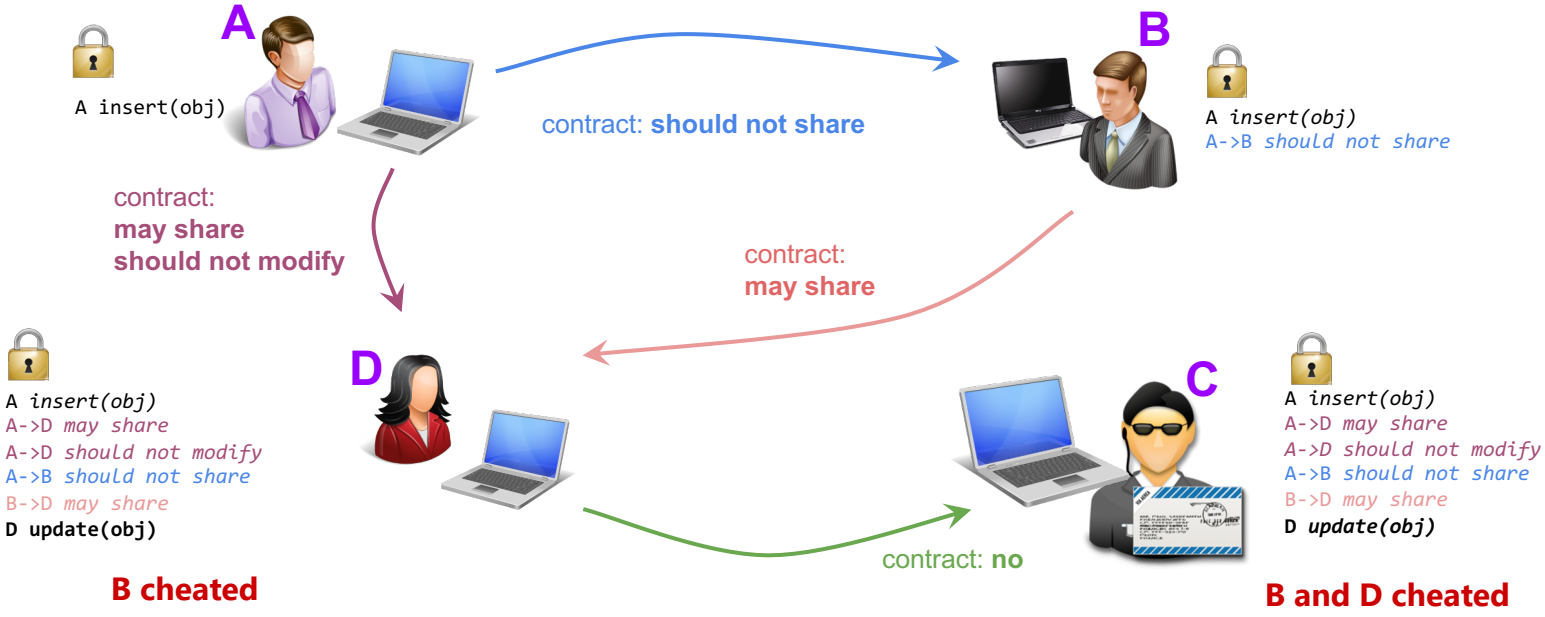
2. Trustworthy Collaboration
   - Evaluate collaborators trust based on their past behavior
   - Evaluation: game theory and user studies

Trustworthy collaboration

- Prediction of future collaborative user behavior based on the quality of user contributions

Trust computation – my contributions

- Respect/Violation of contracts
  - Collaborative editing contracts(share/edit)[TIM12b,TIBM11]



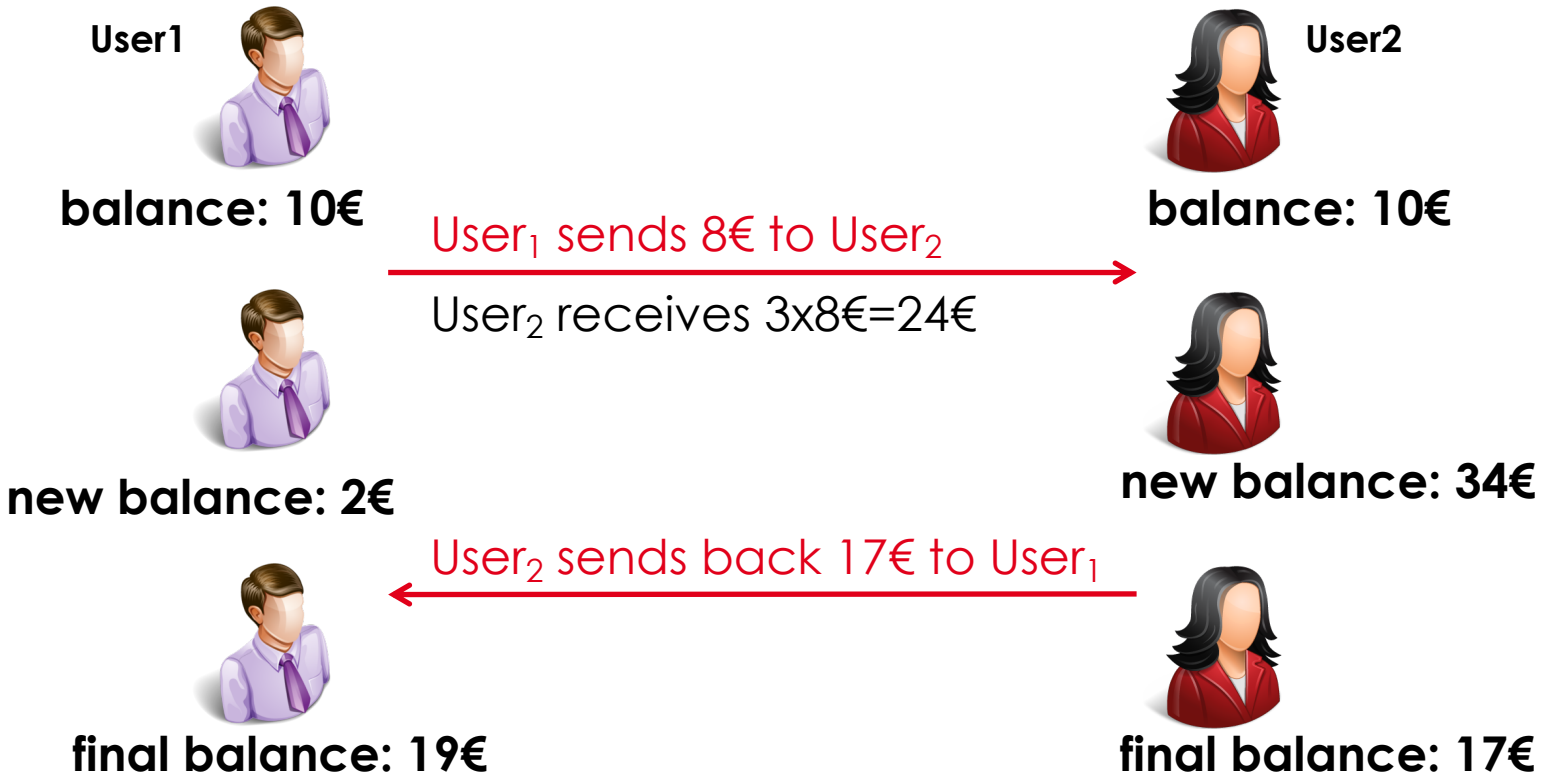
  - Log auditing [TI11,TIM12]
  - Hash-chain based authenticators for log tampering[TIM12a]

Trust computation – my contributions

- Quality assessment of Wikipedia articles
  - Manual feature engineering: extension of ORES (random forest + 11 article features) with readability scores [DI16d]
  - Deep-learning mechanisms: Doc2Vec + DNN [DI16c], RNN-LSTM [DI17]
- Predict trust relations between users that did not interact in the past [DI18]
- Experimental design for testing the proposed trust-based collaboration [IDS19]

Validation of trust-based collaboration

- Using game theory (trust game) [BDM95]



- Collaboration with Department of Psychology Wright State University, Inria USCoast associate team

Trust metric based on user behavior [DI16b]

- For a round:

$$current\_trust_t = \log(send\_proportion_t * (e-1)+1)$$

$$send\_proportion_t = \frac{sending\_amount_t}{\max\_sending\_amount}$$

- For an accumulated number of rounds

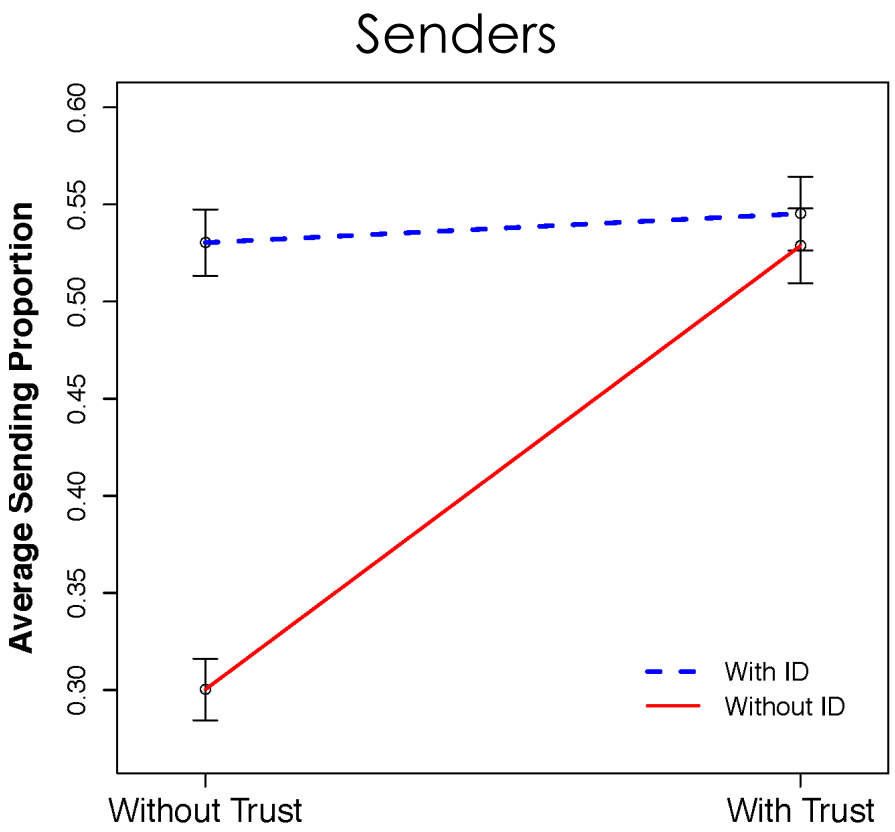
$$aggregate\_trust_t = \alpha_t * current\_trust_t + (1-\alpha_t) * aggregate\_trust_{t-1}$$

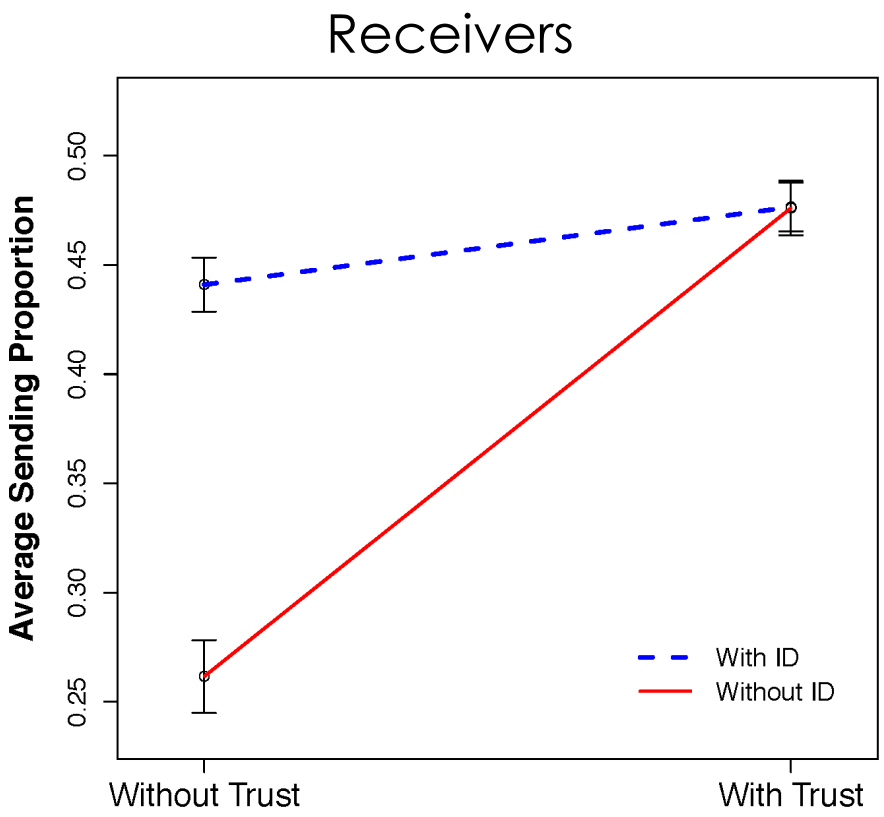
- The trust metric deals with fluctuating user behavior

Experiment design [IDS19]

- 5 sessions with 6 participants each
  - Play 4 trust games
    - Simple Game
    - Identity Game
    - Score Game
    - Combine Game
  - Use z-tree for trust games implementation
  - Each participant received a 10 euro gift card
  - Participant with highest gain received an extra card

Effect of trust and identity





- Showing either trust score or nickname improves the measure
- Showing both of them does not change the measure relative to one of them

Validation of trust-based collaboration - findings

- Trust score or ID availability could significantly improve the level of cooperation between users
- Trust-based systems could be a replacement of identity-based systems especially in the context of large scale collaboration

Trustworthy collaboration - summary

- Contributions towards an **optimistic trust-based security** for large scale peer-to-peer collaboration
  - Modeling of **contract-based collaboration**
  - **Computation** of **users trust** based on their past behavior and **prediction** of their future behavior
  - **Evaluation** using **game theory** and **user studies**

Future directions – Secure and trustworthy distributed collaborative systems

- Security for distributed collaboration without central authority
- Users trust evaluation based on past contributions
- Replication mechanisms for complex data

Security for distributed collaboration without central authority

- Access control without central authority
- End-to-end encryption with group key management

Users trust evaluation based on past contributions

- In Wikipedia
  - Use the proposed trust metric: User contributions throughout article revisions are similar to user interactions in the trust game
  - Need of quality metrics: how long an edit of a user persists

- BPI Deeptech project with Fair&Smart (2020-2023)
  - Computing trust among clients and enterprises for a personal data management platform that respects GDPR

Replication mechanisms for complex data

- Composition of state-based CRDT, operation-based CRDT and operational transformation
- CRDT for relational database [YI20]
- Maintaining global invariants for CRDT such as integrity constraints (uniqueness, reference integrity, numeric constraints) for relational databases

Large-scale trustworthy distributed collaborative systems

- New uses and new practices due to large scale adoption
- New challenges
  - Consistency of complex replicated data in large scale collaboration
  - Large scale group awareness
  - Large scale group user studies
  - Trust and Security without a central authority

List of co-authors

Luc André
Mehdi Ahmed-Nacer
Lamia Benmouffok
Mohamed-Rafik Bouguelia
Jean-Michel Busca
Michèle Cart
François Charoy
Quang-Vinh Dang
Victorien Elvinger
Jean-Philippe Eisenbarth
Jean Ferrié
Olivia Fox
Rachid Guerraoui
Anne-Marie Kermarrec
Stéphane Martin
Pascal Molli
Meagan Newman
Matthieu Nicolas
Moira Norrie
Hoang-Long Nguyen
Hoai-Le Nguyen
Gérald Oster
Stavroula Papadopoulou
Olivier Perrin
Charbel Rahhal
Hyun-Gul Roh
Valerie Shalin
Marc Shapiro
Hala Skaf-Molli
Pierre Sutra
Hien Thi Thu Truong
Pascal Urso
Weihai Yu

Thank You

ANY QUESTIONS?

Claudia Ignat

Habilitation defense
April 23, 2021
claudia.ignat@inria.fr

References

[AMOI13] Supporting Adaptable Granularity of Changes for Massive Scale Collaborative Editing, L. André, S. Martin, G. Oster and C.-L. Ignat. CollaborateCom 2013.

[AIOR11] Evaluating CRDTs for real-time document editing. Mehdi Ahmed-Nacer, Claudia-Lavinia Ignat, Gérald Oster, Hyun-Gul Roh, Pascal Urso. ACM Symposium on Document Engineering 2011

[BDM95] Trust, social history, and reciprocity. J. Berg, J. Dickhaut, and K. McCabe, Games and Econ. Behav.1995.

[EG89] Concurrency Control in GroupWare Systems, C.A. Ellis and S.J. Gibbs. ACM SIGMOD Record 18(2), 1989.

[EGR91] Groupware: Some issues and experiences, C.A. Ellis, S. J. Gibbs and G. Rein, Communications of the ACM,1991

[DI16a] Performance of real-time collaborative editors at large scale: User perspective. Quang Vinh Dang, Claudia-Lavinia Ignat. Networking 2016

[DI16b] Computational Trust Model for Repeated Trust Games. Quang Vinh Dang, Claudia-Lavinia Ignat. Trustcom 2016

[DI16c] Quality Assessment of Wikipedia Articles without Feature Engineering. Quang Vinh Dang, Claudia-Lavinia Ignat. JCDL 2016

[DI16d] Measuring Quality of Collaboratively Edited Documents: The Case of Wikipedia. Quang-Vinh Dang, Claudia-Lavinia Ignat. CIC 2016

References

[DI17] An end-to-end learning solution for assessing the quality of Wikipedia articles. Quang-Vinh Dang, Claudia-Lavinia Ignat. OpenSym 2017

[DI18] Link-Sign Prediction in Dynamic Signed Directed Networks. Quang-Vinh Dang, Claudia-Lavinia Ignat. CIC 2018

[IAO17] Enhancing rich content wikis with real-time collaboration. Claudia-Lavinia Ignat, Luc André, Gérald Oster. Concurr. Comput. Pract. Exp. 2017

[IDS19] The influence of trust score on cooperative behavior. Claudia-Lavinia Ignat, Quang-Vinh Dang, and Valerie L. Shalin. ACM Trans. Internet Technol. 19, 4, Article 46, September 2019

[IN08] Multi-level Editing of Hierarchical Documents. Claudia-Lavinia Ignat, Moira C. Norrie. Comput. Support. Cooperative Work. 2008

[IO07] Flexible Reconciliation of XML Documents in Asynchronous Editing. Claudia-Lavinia Ignat, Gérald Oster. ICEIS 2007

[IO08] Awareness of Concurrent Changes in Distributed Software Development. Claudia-Lavinia Ignat, Gérald Oster. OTM Conferences (1) 2008

[IOMC07] A Comparison of Optimistic Approaches to Collaborative Editing of Wiki Pages. C. Ignat, G. Oster, P. Molli, M. Cart, J. Ferrié, A. Kermarrec, P. Sutra, M. Shapiro, L. Benmouffok, J. Busca, and R. Guerraoui. CollaborateCom 2007.

[IOFSC15] How Do User Groups Cope with Delay in Real-Time Collaborative Note Taking. Claudia-Lavinia Ignat, Gérald Oster, Olivia Fox, Valerie L. Shalin, François Charoy. ECSCW 2015

References

[IPON08] Providing awareness in multi-synchronous collaboration without compromising privacy. Claudia-Lavinia Ignat, Stavroula Papadopoulou, Gérald Oster, Moira C. Norrie. CSCW 2008

[NEOIC17] MUTE: A Peer-to-Peer Web-based Real-time Collaborative Editor. Matthieu Nicolas, Victorien Elvinger, Gérald Oster, Claudia-Lavinia Ignat, François Charoy. ECSCW, 2017

[NI18] An Analysis of Merge Conflicts and Resolutions in Git-Based Open Source Projects. Hoai Le Nguyen, Claudia-Lavinia Ignat. Journal of Computer Supported Cooperative Work, 2018

[NI20] Time-Position Characterization of Conflicts: A Case Study of Collaborative Editing.Hoai Le Nguyen, Claudia-Lavinia Ignat. CollabTech 2020

[SPBZ11] Conflict-free Replicated Data Types, M. Shapiro, N. Preguica, C. Baquero, M. Zawirski. Research Report, RR-7687, INRIA, 2011

[TI11] Hien Thi Thu Truong, Claudia-Lavinia Ignat. Log Auditing for Trust Assessment in Peer-to-Peer Collaboration. ISPDC 2011

[TIBM11] A contract-extended push-pull-clone model. Hien Thi Thu Truong, Claudia-Lavinia Ignat, Mohamed-Rafik Bouguelia, Pascal Molli. CollaborateCom 2011

[TIM12a] Authenticating operation-based history in collaborative systems. Hien Thi Thu Truong, Claudia-Lavinia Ignat, Pascal Molli. GROUP 2012

References

[TIM12b] A Contract-Extended Push-Pull-Clone Model for Multi-Synchronous Collaboration. Hien Thi Thu Truong, Claudia-Lavinia Ignat, Pascal Molli. Int. J. Cooperative Inf. Syst. 2012

[WUM09] Logoot : a Scalable Optimistic Replication Algorithm for Collaborative Editing on P2P Networks, S. Weiss, P. Urso and P. Molli. ICDCS 2009.

[YI15] A CRDT Supporting Selective Undo for Collaborative Text Editing. Weihai Yu, Luc André, Claudia-Lavinia Ignat. DAIS 2015

[YEI19] A Generic Undo Support for State-Based CRDTs. Weihai Yu, Victorien Elvinger, Claudia-Lavinia Ignat. OPODIS 2019

[YI20] Conflict-Free Replicated Relations for Multi-Synchronous Database Management at Edge. Weihai Yu, Claudia-Lavinia Ignat. IEEE International Conference on Smart Data Services, 2020