# Um Olhar Atual sobre Sistemas Distribuídos: Da Pesquisa à Aplicação no Mundo Real

## Workshop Suíça–Brasil @ PUCRS + Online

📍 PUCRS – Escola Politécnica – Prédio 32, Auditório Térreo
📅 15 e 16 de abril de 2025
🔗 Gratuito e aberto — presencial ou online!

Você já se perguntou como blockchain funciona por trás dos panos? Ou como grandes empresas mantêm seus sistemas disponíveis 24/7? A resposta está em uma área fascinante da computação: **Sistemas Distribuídos**. Neste workshop de dois dias, pesquisadores e profissionais do Brasil e da Suíça se reúnem para explorar como projetar sistemas **confiáveis e escaláveis** — desde os fundamentos teóricos até casos reais de sucesso em blockchain, nuvem, bancos e telecom.

### 🧠 **O que você vai descobrir:**

- Como máquinas chegam a um acordo mesmo com falhas (consenso)
- Como funcionam blockchains, smart contracts e armazenamento "cloud-of-clouds"
- Estratégias para escalar sistemas com segurança
- Como transformar pesquisa em startups e soluções reais
- Histórias de carreira: da universidade ao mercado

💡 Não é só sobre algoritmos — é sobre **colaboração, pessoas e futuro**.

### Quem deve participar?

- 👨‍🎓 Estudantes de graduação e pós-graduação
- 👨‍🔬 Jovens pesquisadores interessados em sistemas distribuídos
- 👨‍💻 Desenvolvedores que curtem blockchain, cloud, tolerância a falhas
- 🤝 Qualquer pessoa curiosa sobre tecnologia distribuída

Junte-se a nós para dois dias de conversas técnicas, trocas de experiências e muito aprendizado em um ambiente descontraído e acolhedor. Traga suas dúvidas, ideias e curiosidade!

☕ Sim, tem coffee break.
🎤 Sim, a maioria das palestras será em português.
💻 Sim, dá pra participar online também.
📝 Por favor, preencha o pré-registro online — isso nos ajuda a organizar tudo direitinho.
https://forms.office.com/r/VLG1LBtGSQ
🌎 O link para participar on-line será divulgado em fldotti.github.io
Vamos desvendar juntos os bastidores dos sistemas distribuídos!

📋 Organização :

Fernando Dotti (PUCRS/Brazil)
fldotti.github.io

Fernando Pedone (USI/Suíça)
www.inf.usi.ch/pedone/

💰 Apoio:

FAPERGS
Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

Program
April 15th

13:30 "What Blockchains, Telecoms, and Banks Have in Common –
The Secret Life of Consensus in Distributed Systems"
Fernando Pedone (USI-Switzerland)

Consensus is a fundamental problem in distributed systems, concerned with getting a group of machines to agree on a single value despite failures or delays. Since the seminal work by Lamport, Shostak, and Pease, consensus has become the backbone of many fault-tolerant systems. This talk explores the theory and practice of consensus, beginning with its foundations and examining how it underpins critical replication strategies. We will delve into three real-world case studies the presenter has experienced first-hand: database replication, which uses consensus to maintain consistency across replicas; large-scale graph processing systems, where coordination is essential to ensure consistency across graph shards; and blockchain, focusing on modern consensus protocols like Proof of Stake. Finally, we will discuss emerging directions for consensus and the exciting opportunities awaiting those who master the art and science of distributed agreement.

Bio: **Fernando Pedone** is a Professor of Computer Science at the Faculty of Informatics, Università della Svizzera italiana (USI) in Lugano, Switzerland. His research focuses on distributed systems, particularly fault tolerance, consistency, and scalable replication. Over the years, he has contributed to the theoretical foundations and practical implementations of consensus protocols, and collaborated with industry and academia to bring cutting-edge distributed technologies to life. He also maintains a long-standing research collaboration with Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS) in Brazil. In this talk, he brings insights from real-world systems and decades of research to demystify the critical role of consensus in modern computing.

14:00 "Quase 100 anos de cooperação em computação distribuída -
Transcorrer, Resultados e Perspectivas"
Fernando Dotti (PUCRS)

Esta apresentação contextualiza os assuntos e contribuições de pesquisa desenvolvidos no escopo da colaboração entre apresentadores deste workshop. Isto é feito mostrando tanto as relações entre os autores envolvidos como sumarizando os desafios e resultados enfrentados.

Bio: **Fernando Dotti** obteve seu doutorado na Universidade Técnica de Berlim em 1997 e desde 1998 é professor na PUCRS atuando como membro permanente do PPGCC. Seus assuntos de pesquisa giram em torno da construção de sistemas distribuídos confiáveis, onde contribui desde 2013 com diversas abordagens para escalabilidade de da abordagem de Replicação Máquina de Estados. Antes disso contribuiu com abordages e experiências para especificação e verificação de algoritmos distribuídos. Ele mantém colaboração ativa com a USI/Suíça e parceiros brasileiros como UnB, UFU, UFSC.

14:30 "Uma Carreira Consistente em Computação Distribuída"
Lásaro Camargos

A computação distribuída é a base que torna viáveis muitas das inovações tecnológicas que dominam a mídia, como inteligência artificial, blockchain, jogos online, computação em nuvem e plataformas de streaming. Essas tecnologias exigem arquiteturas distribuídas com alto grau de escalabilidade,

confiabilidade e corretude—características que, muitas vezes, entram em conflito. Como profissionais da área, precisamos constantemente encontrar o melhor equilíbrio entre esses fatores, o que torna nosso trabalho desafiador e estimulante. Nesta apresentação, compartilharei minha trajetória como pesquisador e desenvolvedor em computação distribuída, discutindo os desafios e oportunidades dessa carreira essencial, mas frequentemente subestimada.

Bio: **Lásaro** já foi aluno e professor, trabalhou tanto em gigantes da indústria quanto em startups anônimas. Escreveu código em Python, C, Java, C#, Go, Rust e até JavaScript—sem pre buscando novos desafios. E consistentemente encontra na computação distribuída os mais interessantes.

15:00 Coffee-break

15:30 "Atomic Multicast: do Skeen ao Pacheco"
Paulo Coelho (UFU)

Construir aplicações distribuídas é uma tarefa complexa, e o particionamento de dados, ao mesmo tempo que possibilita aumentar a escalabilidade das aplicações, introduz complexidade adicional ao coordenar operações entre grupos de réplicas. O Multicast atômico é uma abstração de comunicação que simplifica o raciocínio e a construção de sistemas particionados. Ele permite que mensagens sejam entregues de forma confiável a um subconjunto dos grupos do sistema, garantindo ao mesmo tempo uma ordem parcial nas mensagens entregues. Esta apresentação apresenta um pouco do histórico do desenvolvimento de algoritmos de multicast atômico baseados em timestamp, enfatizando a importância de reduzir o número de passos de comunicação para melhor o desempenho de sistemas modernos geograficamente distribuídos.

Bio: **Paulo Coelho** é professor associado na Faculdade de Computação da Universidade Federal de Uberlândia, possui mestrado pela Universidade de Campinas (UNICAMP) e doutorado pela Universidade de Lugano (USI), na Suíça. Suas áreas de pesquisa incluem tolerância a falhas e comunicação em grupo.

16:00 "Composing State Machine Replicas"
Odorico Mendizabal (UFSC)

State Machine Replication (SMR) is a fundamental technique for building highly available and fault-tolerant services. While SMR has been extensively studied and optimized over the years, one aspect remains unexplored: service composition. This talk introduces a new perspective on SMR by defining its compositionality and demonstrating how combining replicated state machines can lead to more modular, scalable, and flexible systems. By combining SMR through composition, we enable the reuse of existing SMRs to construct complex services while maintaining fault tolerance and consistency. This approach aligns with modern large-scale applications, particularly in cloud computing and microservices architectures. Composition enhances SMR by allowing new functionalities to be added to existing replicas and enabling different replicas to collaboratively execute parts of the same operation. Additionally, it facilitates sharding and state partitioning, where disjoint state variables can be assigned to separate SMRs, improving efficiency. Throughout the talk, we will explore illustrative examples showcasing the benefits of SMR composition and discuss open challenges in this evolving research direction.

Bio: **Odorico Machado Mendizabal** is professor at the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil. He holds an M.Sc. and Ph.D. in Computer Science from the Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil. He is interested in the theory and practice of dependable systems, developing research in fault tolerance, reliable communication protocols, and replication and recovery of distributed systems.

16:30 "A different form of consensus: Proof of Work & Proof of Stake"
Enrique Fynn (Chorus One)

This takl explores consensus mechanisms in blockchains and their role in ensuring decentralized trust. A brief overview of Proof of Work (PoW) and Proof of Stake (PoS) highlighting their strengths and trade-offs. Smart contracts: How they enable secure, automated, and tamper-proof agreements on the blockchain. Finally, I will share my practical experience developing smart contracts and key challenges faced during implementation. The session will conclude with insights gained and best practices for building smart contract-based applications

Bio: **Enrique Fynn** graduou-se em Ciência da Computação e concluiu o mestrado na UFU (Universidade Federal de Uberlândia), fez doutorado em Sistemas Distribuídos na USI (Università della Svizzera Italiana). Sob a supervisão do professor Fernando Pedone, desenvolveu e defendeu a tese intitulada "Scaling Blockchains". Há 4 anos, trabalha na Chorus One, uma empresa que oferece infraestrutura e staking em blockchains como desenvolvedor.

17:00 "Byzantine Consensus and Semantic Gossip: Scaling Decentralized Systems."
Ricardo Guimarães (PUCRS)

Gossip allows to scale communication to larger sets of nodes and has been recently considered as a communication layer to support consensus. Consensus protocols tend to have high redundancy regarding the number of messages exchanged between participating machines, as the algorithm's termination requires only a subset of machines to reach an agreement. Similarly, epidemic communication protocols, such as Gossip, introduce even more redundancy by allowing the same machine to receive the same message from different peers. By their nature, both consensus protocols and gossip handle message losses and process failures. This double redundancy means that it is not yet clear how efficient it is to use gossip as a black building-box for consensus. We discuss how far redundancy can be eliminated without compromising consensus' safety and liveness and propose to selectively eliminate redundancy considering the combined use of protocols.

Bio: **Ricardo Guimarães** is a Master's student in Computer Science at the Pontifical Catholic University of Rio Grande do Sul and Bachelor's degree in Philosophy from the Federal University of Rio Grande do Sul. Software developer specializing in financial applications for over four years. Main interests in computing: blockchains, Byzantine consensus, and theory of computation.

April 16th

9:00 "Cloud-of-Clouds Storage: from Theory to Production"
Alysson Bessani (Universidade de Lisboa)

Public clouds have become a fundamental pillar for many companies and internet-scale services. In particular, the exponential rise in the amount generated data made (public) cloud storage services a key infrastructure for dealing with such data. Despite the huge efforts of cloud providers to secure their services, there is a profusion of events related to intrusions, unavailability, and even data corruption/destruction. One way to minimize the effect of such events is to exploit the existence of a vibrant cloud market and store data on a set of different providers, implementing a cloud-of-clouds storage strategy. To do that, cloud users must rely on fault-tolerant storage abstractions built on top of fail-prone individual cloud services, namely read/write registers and lease objects. Many proposals for such abstractions come from both the theoretical and the systems communities. Based on a subset of these works, we select and present four constructions (three types of registers and one lease object), comparing their properties and costs, and discuss how they were used to implement practical cloud-of-clouds storage services (from a programming library to a commercial storage service, vawlt.io). In the end, we'll also share some of our experience in transforming fundamental research done in academia in a startup that secured more than 3M euros of VC funding.

Bio: **Alysson Bessani** is a Professor at the University of Lisbon Faculty of Sciences, Portugal, and the director of the LASIGE research unit. He received his Ph.D. in Electrical Engineering from UFSC (Brazil) in 2006, was a visiting professor at Carnegie Mellon University (2010), and a visiting researcher at Microsoft Research Cambridge (2014). Alysson coordinated/collaborated on 13 international projects and co-authored more than 120 peer-reviewed publications on dependability, security, Byzantine fault tolerance, and cloud storage. He is also the principal researcher behind the BFT-SMaRt consensus library (http://bft-smart.github.io/library/) and a co-founder of the Vawlt dependable & secure cloud storage startup (https://vawlt.io). More information about him can be found at http://www.di.fc.ul.pt/~bessani.


9:30 "How Far Can Synchronous BFT Consensus Go?"
Nenad Milošević (USI - Informal Systems)

Synchronous BFT consensus protocols offer strong fault tolerance but are often met with skepticism and viewed as impractical due to their reliance on strict timing bounds. In this talk, we present a fresh perspective on this problem and share our recent efforts toward designing more practical synchronous protocols. We start by analyzing how synchrony violations can compromise protocol correctness, even in the presence of malicious behavior. Using the BoundBFT protocol as a case study, we show that communication redundancy allows some synchronous protocols to tolerate a degree of synchrony violations without breaking safety. These results motivated a three-month empirical study of message delays in real-world networks. Based on the findings, we introduced a hybrid synchrony model that captures the observed latency patterns by distinguishing between small and large messages. Within this model, we designed AlterBFT, a new protocol that retains the fault tolerance of synchronous protocols while offering significantly better performance. This work represents a first step toward reassessing the practicality of synchronous consensus, and we plan to further explore their real-world applicability.

Bio: **Nenad Milošević** is a research engineer at Informal Systems and a scientific collaborator at Università della Svizzera italiana (USI). He recently completed his PhD at USI, where his research focused on distributed consensus protocols, including the interplay between gossip and consensus, the robustness of synchronous BFT algorithms, and how to design

consensus protocols that achieve both high performance and strong resilience under realistic network conditions. At Informal Systems, Nenad works on Malachite, a Rust-based implementation of the Tendermint consensus algorithm, where he is involved in both protocol design and software development. His work bridges theory and practice, and his research has been published at well-established venues in the distributed systems community, including Middleware, SRDS, and OPODIS.

10:00 "Consenso Bizantino Baseado em uma Camada de Rede com Ordenação de Mensagens Tolerante a Intrusões"
Eduardo Alchieri (UnB)

Trabalhos recentes propuseram protocolos de consenso que dividem as tarefas de garantir as propriedades de acordo e de terminação entre as camadas de rede e de aplicação, respectivamente. A camada de rede precisa entregar as mensagens já ordenadas para garantir acordo. Para tanto, essas soluções normalmente utilizam um sequenciador. No entanto, tolerar um sequenciador malicioso traz novos desafios, uma vez que um mesmo número de sequência pode ser atribuído para diferentes mensagens e enviadas para diferentes réplicas. Para contornar tais problemas, o NeoBFT emprega um passo adicional de comunicação entre as réplicas. Apesar de lidar com o problema, essa abordagem impacta o desempenho do sistema, uma vez que sincronizações adicionais são necessárias nas réplicas antes da execução das requisições. Este trabalho propõe o NsoBFT (Network secure ordered BFT), um protocolo de consenso que utiliza uma camada de rede com um serviço seguro de ordenação de mensagens implementado através do componente seguro USIG (Unique Sequential Identifier Generator). Desta forma, não é necessário executar sincronizações adicionais nas réplicas antes de executar as requisições. Resultados experimentais comparam as propostas similares encontradas na literatura e mostram que esta característica faz com que o NsoBFT tenha um desempenho superior ao NeoBFT.

*English version: recent work proposes consensus protocols that divide the tasks of ensuring agreement and termination properties between the network and application layers, respectively. The network layer must deliver ordered messages to ensure agreement. To achieve this, these solutions use a sequencer. However, tolerating a malicious sequencer brings new challenges, since it can assign the same sequence number to different messages and send them to different replicas. To circumvent this problem, NeoBFT employs an additional communication step between replicas. Although this approach mitigates the problem, it affects system performance by requiring additional synchronization between replicas before requests are executed. This work proposes NsoBFT (Network Secure Ordered BFT), a consensus protocol that uses a network layer with a secure message ordering service implemented by the secure component USIG (Unique Sequential Identifier Generator). Thus, it is not necessary to perform additional synchronizations among the replicas before executing requests. Experimental results compare NsoBFT with related work, and show that this feature allows NsoBFT to surpass the performance of NeoBFT.*

Bio: **Eduardo Alchieri** recebeu os títulos de Bacharel em Ciência da Computação (INE), Mestre em Engenharia Elétrica (PPGEEL) e Doutor em Engenharia de Automação e Sistemas (PPGEAS) pela Universidade Federal de Santa Catarina (UFSC). Atualmente éprofessor do Departamento de Ciência da Computação (CIC) da Universidade de Brasília (UnB). Seus interesses de pesquisa incluem aspectos teóricos e práticos sobre sistemas distribuídos, com especial interesse nas áreas de tolerância a falhas e intrusões e de segurança computacional.

10:30 Coffee Break

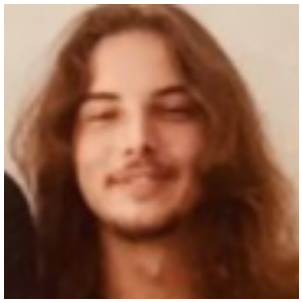
11:00 "SkipLists in SMR: Simplifying and Enhancing State Transfer and Validation"
Eliã Batista (USI - PUCRS)

State machine replication (SMR) ensures fault tolerance in distributed systems by having replicas execute requests consistently. A key aspect of SMR performance is replica state management.
State management is crucial for ensuring durability and consistent state even if some replicas fail or behave arbitrarily. We investigate the use of optimized data structures to cluster data and enable parallel state transfer and separate cluster validation. In particular, we investigate the use of SkipLists as self-validating clustered data structures. We propose a novel SkipLists-based data structure that supports

both self-validation and efficient data clusterization. SkipLists offer some advantages over the commonly used Merkle trees. While Merkle trees provide cryptographic proofs of integrity, their hierarchical balanced structure leads to overhead and complex updates. In contrast, SkipLists offer similar performance for insertion and search operations while keeping simplicity with probabilistic balancing. Additionally, the layered design of SkipLists naturally supports clusterization and cluster validation without requiring full traversals. By leveraging these properties, our SkipLists-based approach aims to enhance the efficiency and scalability of SMR systems, improving fault tolerance and recovery performance.

Bio: **Eliã Batista** is a PhD researcher in the Distributed Systems Group at Università della Svizzera italiana (USI), Lugano, Switzerland, under a cotutelle de thèse program with PUC-RS, Brazil. His research focuses on distributed dependable systems, including State Machine Replication (SMR), parallel scheduling, state management, and recovery. He is currently working on optimized data structures for SMR state transference and Atomic Multicast algorithms. He holds a M.Sc. and B.Sc. in Computer Science from PUC-RS and has published in leading journals and conferences in the Distributed Computing Community. Before academia, he worked for over a decade as a senior software developer and system analyst.

11:15 "Performance Evaluation of a Consensus Algorithm"
Lucca Dornelles Cezar (PUCRS)



This talk reports on the current experiences and results towards evaluating the performance of the Tendermint consensus protocol, which is being carried out in the scope of a collaboration among PUCRS and Informal Systems.

Bio: **Lucca Dornelles Cezar** is a Master's student in Computer Science at PPGCC-PUCRS. Graduated in Computer Science, during the course he studied infrastructure and linear programming until he wrote the final project on consensus algorithms. He is currently studying the performance of consensus algorithms.

# Um Olhar Atual sobre Sistemas Distribuídos: Da Pesquisa à Aplicação no Mundo Real

Workshop Suíça–Brasil @ PUCRS + Online

📍 PUCRS – Escola Politécnica – Prédio 32, Auditório Térreo

📅 15 e 16 de abril de 2025

🔗 Gratuito e aberto — presencial ou online!

Program
April 15th

13:30 "What Blockchains, Telecoms, and Banks Have in Common – The Secret Life of Consensus in Distributed Systems"
Fernando Pedone (USI-Switzerland)

14:00 "Quase 100 anos de cooperação em computação distribuída - Transcorrer, Resultados e Perspectivas"
Fernando Dotti (PUCRS)

14:30 "Uma Carreira Consistente em Computação Distribuída"
Lasaro Camargos

15:00 Coffee-break

15:30 "Atomic Multicast: do Skeen ao Pacheco"
Paulo Coelho (UFU)

16:00 "Composing State Machine Replicas"
Odorico Mendizabal (UFSC)

16:30 "A different form of consensus: Proof of Work & Proof of Stake"
Enrique Fynn (Chorus One)

17:00 "Byzantine Consensus and Semantic Gossip: Scaling Decentralized Systems."
Ricardo Guimarães (PUCRS)

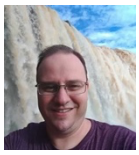
17:15 -

April 16th

9:00 "Cloud-of-Clouds Storage: from Theory to Production"
Alysson Bessani

9:30 "How Far Can Synchronous BFT Consensus Go?"
Nenad Milošević (USI - Informal Systems)

10:00 "Consenso Bizantino Baseado em uma Camada de Rede com Ordenação de Mensagens Tolerante a Intrusões"
Eduardo Alchieri (UFSC)



10:30 Coffee Break

11:00 "SkipLists in SMR: Simplifying and Enhancing State Transfer and Validation"
Eliã Batista (USI / PUCRS)

11:15 "Performance Evaluation of a Consensus Algorithm"
Lucca Dornelles Cezar (PUCRS)

11:30 Closing