

# 徐亮亮

✉ llxu@mail.ustc.edu.cn · ☎ (+86) 15656547361 · 🌐 <https://lianglxu.github.io>

## 个人介绍

我的主要研究方向是分布式存储系统、数据修复、纠删码以及新型内存分离的存储架构。博士期间，我的工作主要包括设计有效的数据布局，高效的故障修复算法，设计实际的纠删码部署策略，最后在分布式存储系统中实现。相关论文发表至 IEEE INFOCOM, IEEE IPDPS, USENIX HotStorage, IEEE TIT 以及 IEEE TPDS 等国际顶级会议及著名期刊上。其中，CCF 推荐列表 A 类论文共发表 5 篇，以第一作者（或学生一作）身份发表论文共 6 篇。

## 教育背景

博士，中国科学技术大学，合肥

2017 – 至今 (预计 2022.06 毕业)

- 计算机科学与技术，先进数据系统实验室
- 导师：许胤龙 & 吕敏

本科，安徽大学，合肥

2013 – 2017

- 信息与计算科学
- GPA: 3.69/4.0

## 实习经历

华为技术有限公司，云存储 lab，内存存储项目组，深圳

2020 年 10 月 – 2021 年 01 月

- 导师: 左鹏飞
- 实习内容：纠删码在内存存储系统中的应用，具体如下：
  - 调研了以下几个方面工作：EC 在 cache 场景下的应用、RDMA 环境下的高性能 key value 数据库、RDMA 的一些高级硬件结合级操作（如 doorbell）、EC 在 x86 架构下的应用最多的两个库（Jerasure 和 ISA-L）的设计文档与实现、EC 在 data-intensive 场景下的应用（如 memcached）以及 Disaggregating Persistent Memory 等场景。
  - 测试了大量 Jerasure 和 ISA-L 库的性能。如：小对象编解码 latency、throughput，多线程编解码性能，预加载 SSE 指令集的编解码性能，EC 细粒度参数 word、packet、buffer size 等灵敏度性能，pipeline 编解码性能，不同 EC 编码类型的性能（RS、CRS 等）。
  - 发现了一些现象并分析了性能瓶颈：如采用预加载指令集后，EC 在小对象时延较低（<8k 对象在 1us 以下）；EC 编码对象时延的增加比太大，甚至大于线性增长，主要原因可能是大对象计算编码时 L1 和 L3 等 cache miss 快速增加；单边的 RDMA 读写性能在小对象时延高于 EC 编码，但 RDMA 读写时延增加缓慢；EC 编码与 RDMA 写之间的 latency 存在 GAP，小对象 EC 编解码时延较低，但大对象 RDMA 读写时延较低。
  - 提出了一些解决方案：考虑 in-memory KV 数据库场景：如关于写性能使用动态 pipeline 编解码，先把大对象 split 为多个小对象，串行的编解码性能更好，这样能够把编码时延降下来，接着再动态的合并小 split（基于网卡端 QP 的负载以及 CQE 状态），进一步优化网络消息转发；关于读性能长尾延迟优化，可采用多读校验块、设计低复杂度生成矩阵（一个 parity 计算简单），reorder 到达的消息来避免不必要的降级读；多线程来优化降级读解码开销等策略。

## 项目

1. **PDL**. This project proposes an efficient PBD-based (Pairwise Block Design) Data Layout, PDL, to speed up data repair for single node failure in mixed erasure-coded distributed storage systems. It achieves almost uniform distribution, and higher repair performance due to reduced cross-rack traffic and load balance of read and write I/Os during repair process. I design the data distribution method, and the corresponding failure recovery scheme. And I also implement them in Hadoop 3.1.1.
2. **SelectiveEC**. This project proposes a balanced scheduling module, SelectiveEC, to dynamically select some stripes to be reconstructed in a batch, and select source and replacement nodes for each reconstruction task. It achieves balanced network recovery traffic, computing resources and disk I/Os against single node failure

- in erasure-coded storage systems. I design the scheduling algorithm, build the SelectiveEC prototype and validate it by simulation.
3.  $D^3$ . The proposed distribution  $D^3$  uniformly distributes data/parity blocks among nodes in large scale erasure-coded distributed storage systems, and minimizes the cross-rack repair traffic against a single node failure. I integrate the distribution  $D^3$  into HDFS-EC module of Hadoop 3.1.0 and evaluate the repair performance over Reed-Solomon codes. In the journal version, I extend it to locally repairable codes, provide efficient strategy to maintain the  $D^3$  data layout after recovery, and conduct more experimental evaluations.
  4. **A note on one weight and two weight projective  $Z_4$ -codes.** This is the work in my undergraduate. I solve the open problems about algebraic codes, moreover, I work out the diophantine problem and then give the sufficient conditions for the nonexistence of two-Lee weight projective codes over  $Z_4$  with type  $4^{k_1}2^{k_2}$ .

## 发表论文

---

1. SelectiveEC: Selective Recovery in Erasure-coded Storage Systems.  
**Liangliang Xu**, Min Lyu, Qiliang Li, Lingjiang Xie, Yinlong Xu and Cheng Li.  
Submitted to 2021 USENIX Annual Technical Conference (USENIX ATC 2021).
2. A Data Layout and Fast Failure Recovery Scheme for Distributed Storage Systems with Mixed Erasure Codes.  
**Liangliang Xu**, Min Lyu, Zhipeng Li, Cheng Li and Yinlong Xu.  
Submitted to IEEE Transactions on Computers (TC 2021).
3. PDL: A Data Layout towards Fast Failure Recovery for Erasure-coded Distributed Storage Systems.  
**Liangliang Xu**, Min Lv, Zhipeng Li, Cheng Li and Yinlong Xu.  
IEEE International Conference on Computer Communications (INFOCOM 2020) accepted.  
(AR: 268/1354 = 19.8%, CCF A)
4. Deterministic Data Distribution for Efficient Recovery in Erasure-Coded Distributed Storage Systems.  
**Liangliang Xu**, Min Lyu, Zhipeng Li, Yongkun Li and Yinlong Xu.  
IEEE Transactions on Parallel and Distributed Systems (TPDS 2020), 31.10: 2248-2262.  
(CCF A)
5. SelectiveEC: Selective Reconstruction in Erasure-coded Storage Systems.  
**Liangliang Xu**, Min Lyu, Qiliang Li, Lingjiang Xie and Yinlong Xu.  
12th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 2020) accepted.  
(AR: 26/64 = 40.6%)
6. D3: Deterministic Data Distribution for Efficient Data Reconstruction in Erasure-Coded Distributed Storage Systems.  
Zhipeng Li, Min Lv, Yinlong Xu, Yongkun Li and **Liangliang Xu**.  
33rd IEEE International Parallel & Distributed Processing Symposium (IPDPS 2019).  
(AR: 102/372 = 27.7%, CCF B)
7. A note on one weight and two weight projective  $Z_4$ -codes.  
Minjia Shi, **Liangliang Xu** and Gang Yang.  
IEEE Transactions on Information Theory (TIT 2017), 63.1: 177-182.  
(CCF A)

## 专利

---

1. 一种基于纠删码存储系统的负载均衡修复调度方法  
吕敏, 徐亮亮, 李启亮, 谢灵江, 许胤龙  
专利号: 202010313968.5, 申请时间: 2020.04.20, 公开时间: 2020.08.07

## 邀请报告

---

- 2020.07: Paper Presentation in INFOCOM 2020, PDL: A Data Layout towards Fast Failure Recovery for Erasure-coded Distributed Storage Systems, Online.
- 2020.07: Paper Presentation in HotStorage 2020, SelectiveEC: Selective Reconstruction in Erasure-coded Storage Systems, Online.
- 2020.06: Invited Talk in the 18th ChinaSys workshop, PDL: A Data Layout towards Fast Failure Recovery for Erasure-coded Distributed Storage Systems, Online.

## 专业技能

---

- 编程语言: Java. C/C++. Matlab. Python. Linux Shell.
- 分布式系统: HDFS. Ceph. RAMCloud.

## 获奖情况

---

- 深交所奖学金 (¥12000), 2020.
- INFOCOM Student Conference Award (\$ 400), 2020.
- 本科品学兼优毕业生, 2017.
- 优秀本科毕业论文, 2017.
- 本科学术科技一等奖学金, 2016.
- 国际大学生数学建模比赛一等奖, 2016.
- 校级三好学生, 2016.
- 国家励志奖学金, 2014/2015/2016.
- 团学奖学金, 2015.
- 挑战杯校级二等奖, 2014.