

关于编译器的调研报告

李桢弛 22074107

一、开源编译器发展现状与代表项目

1.1 国际主流开源编译器

GCC (GNU Compiler Collection)

提到编译器，我们最先想到的莫过于 GCC。作为最具影响力的开源编译器集合，其广泛支持 C、C++、Fortran 等语言，作为 Linux 生态的核心工具链，具有强大的优化能力和跨平台特性。其拥有广泛的平台支持，交叉编译能力强，适配新硬件架构速度快。其还具有完善的生态系统，GNU 工具链的核心组件，大量第三方工具支持，活跃的开发社区等等。但是，它的代码结构稍显复杂，优化能力落后于 LLVM。

LLVM/Clang

LLVM 最大的特点是模块化设计，且易于扩展，被广泛用于研究和开发。它采用三段式设计，包括前端、优化器和后端，各个组件之间通过统一的中间表示 (LLVM IR) 进行交互，这种架构使得开发新语言或针对新平台变得更加容易。近年来，得益于苹果、微软等科技巨头的深度参与，LLVM 已经成为许多现代编程语言和 AI 框架 (如 Swift、Rust、TensorFlow) 的首选编译工具。其中，Clang 静态分析工具在安全领域发挥重要作用，被广泛用于检测 Linux 内核等重要软件的潜在漏洞。

OpenJDK

OpenJDK 是 Java 平台的官方开源实现，其核心包含了强大的即时编译器 (JIT) 系统。其 HotSpot 虚拟机实现了先进的运行时优化，包括方法内联、逃逸分析、循环优化等技术，同时配备了高效的垃圾回收机制。OpenJDK 具有优秀的平台兼容性，支持跨平台部署，并能与本地代码无缝集成。它在企业级应用开发、大数据处理系统、云计算平台和 Android 应用开发等领域有着广泛应用，已经成为 Java 生态系统的基石。近年来，随着 GraalVM 的发展，OpenJDK 还在探索新的即时编译技术，以提供更好的性能和多语言支持能力。

1.2 中国开源编译器实践

近年来，随着国产化进程的推进，中国在编译器领域也取得了显著进展。特别是在特定领域和场景下，已经形成了具有竞争力的解决方案。目前最具代表性的是华为方舟编译器和腾讯 OpenKona JDK，这两个项目分别在移动终端和服务端领域展现出独特优势。

方舟编译器是华为专门为鸿蒙操作系统开发的编译器平台，其最大特点是支持多语言统一编译。这种创新的技术架构实现了 Java、C++、C 等多种语言的统一编译，简化了开发流程，采用统一的中间代码表示，提高了跨语言优化效率，支持跨语言的联合调试，改善了开发体验。在性能方面，方舟编译器应用启动速度提升超过 30%，运行时内存占用减少约 20%，并通过智能化指令调度和优化，实现性能最大化。在生态建设上，其使用范围主要局限于华为终端设备，开源社区参与度不够活跃，第三方开发者工具支持有限，需要进一步完善文档和开发者资源。

OpenKona JDK 是腾讯针对国产 CPU 架构优化的 Java 开发工具包，专门适配鲲鹏、飞腾等国产处理器架构，优化 JIT 编译器以提升性能，增强了垃圾回收机制，改进了线程调度算法。在云计算场景中表现突出，服务器端应用性能显著提升，大规模分布式系统运行更稳定。在微服务架构下响应更快，资源利用效率更高。在推动 Java 生态国产化方面发挥重要作用，促进了 Java 应用在国产硬件上的适配，提供了完整的开发工具链，形成了活跃的技术社区。

说实话，看到华为方舟编译器和腾讯 OpenKona JDK 这些国产项目，我的内心非常骄傲。以前总觉得编译器这种核心技术是国外的天下，像 GCC、LLVM 这些大佬级项目遥不可及。但现在华为和腾讯能在自己的领域搞出名堂，特别是方舟编译器还能优化鸿蒙系统，让应用跑得更快，这让我觉得咱们中国人也不是只能跟在别人后面跑。不过也得承认，跟国际顶尖水平比，咱们的生态和社区活跃度还有差距，显得有点“自家人玩”。我觉得以后应该多邀请国内外开发者广泛参与，把文档和工具做得更友好，才能真正发展好中国自己的编译器，在国际舞台上占据一席之地。

二、中国产、学、研单位投入编译器研究的必要性分析

掌握编译器核心技术直接关系到中国国家信息基础设施的安全性和可控性。所以，中国开发和研究自己的编译器非常重要。

首先，在一些关键领域如军事、金融、能源等领域的软件开发必须确保安全可控，因为编译过程中可能涉及敏感数据的处理和保护。其次近年来国际形势复杂，技术封锁风险增加，还可能面临高科技领域的出口管制，许可证限制，比如有些商业软件使用许可被限制或撤销。过度依赖国外编译器技术存在潜在风险，包括但不限于核心算法和实现被国外垄断，版本更新和安全补丁依赖国外厂商，关键技术专利和知识产权限制等等。

中国产、学、研单位在编译器领域已经投入颇多，并呈现出不同的特点和侧重。在产业界，以华为、腾讯为代表的科技企业重点投入特定场景的编译器研发：华为方舟编译器针对鸿蒙生态系统优化，支持多语言统一编译，显著提升应用性能；腾讯 OpenKona JDK 则专注于国产 CPU 适配和云计算场景优化。在学术界，以清华大学、中科院计算所等为代表的高校和研究所持续推进编译技术基础研究，建立了多个编译技术实验室，在程序分析、编译优化、并行计算等方向多次取得突破性进展，并培养了大量编译领域人才。在研究机构方面，

以中国电子技术标准化研究院、软件所等为代表的单位主要承担国家重大专项研究，专注于编译器安全性验证、国产平台适配等关键技术攻关，并推动编译技术标准化工作。

我认为，编译器研究的必要性不仅体现在技术层面，更关乎国家战略安全。当前国际技术环境的不确定性，如技术封锁和高科技出口限制，凸显了自主研发的重要性。编译器作为软件开发的核心环节，直接影响关键领域的安全性和国产硬件的生态构建。以华为和腾讯为代表的产业实践，以及高校和科研机构的理论探索，形成了协同发展的格局。我相信，这种模式能够逐步弥补技术差距，为中国信息基础设施的可控性提供保障，同时也为国产软硬件生态的繁荣奠定基础。

三、编译器的安全风险与挑战

编译器的安全风险首先体现在代码生成层面。编译器在进行代码优化时可能引入意外行为，例如优化过程中的重排序可能破坏原有的同步语义，导致并发程序出错；激进的优化策略可能消除看似无用但实际上与安全相关的代码，如内存清零操作的移除可能导致敏感信息泄露。更严重的是，编译器本身可能被植入后门，通过特定的代码模式触发恶意行为，著名的"Trusting Trust"攻击就展示了这种风险。

供应链安全是另一个重要挑战。现代编译器往往依赖大量第三方库和工具，这些依赖项的安全漏洞可能传导至编译器本身。例如，构建工具链中的漏洞可能被攻击者利用来篡改编译过程，插入恶意代码。2023 年 WebAssembly 编译器 Wasmtime 曝出权限提升漏洞。体现了开源编译器依赖社区维护，恶意代码可能混入上游仓库。

编译器优化与安全性的平衡也面临挑战。现代编译器普遍采用复杂的优化策略来提升性能，但这些优化可能与程序的安全假设相冲突。例如，编译器的死代码消除可能移除安全检查代码，内存访问重排可能绕过访问控制检查。特别是在处理加密相关代码时，编译器优化可能无意中创建时序侧信道，导致密钥信息泄露。这要求编译器在优化时需要考虑安全属性的保持，但这往往会影响优化效果。比如编译器激进优化可能导致内存越界或数据竞争。

此外，攻击者可能利用编译器的某些特性或漏洞，构造特殊的输入代码来触发编译器的异常行为，如栈溢出、堆损坏等。更隐蔽的是，攻击者可能利用编译器优化规则的盲点，构造在源码层面看似安全但编译后会产生漏洞的代码，这类攻击特别难以检测。随着 AI 技术在编译器中的应用增加，新的安全风险也随之出现，如 AI 模型可能被污染导致生成有害代码，或被用来自动发现和利用编译器漏洞。正是由于存在如此多的安全风险，中国研制自己的编译技术显得尤为重要。

在调研编译器安全风险时，我深刻认识到其复杂性和潜在威胁。例如，“Trusting Trust”攻击揭示了编译器被恶意篡改的可能性，这让我意识到技术自主的重要性。此外，随着 AI 技术在编译器中的应用，新的安全隐患也随之而来。我认为，国内编译器研发需将安全性置于核心地位，不仅要提升性能，还要在设计之初就防范潜在漏洞，以避免技术成果被外部威

胁利用。

四、编译技术在人工智能发展中的关键作用

首先，先进的编译技术可以加速 AI 模型部署。硬件指令优化，案例：LLVM 的 MLIR (Multi-Level IR) 框架为 TPU/NPU 设计专用中间表示，提升 TensorFlow 模型推理速度 40%。还可以提供异构计算支持，比如场景：编译器将 AI 模型分割为 CPU/GPU/FPGA 可执行模块，降低延迟（如英伟达 CUDA 编译器）。

反过来，AI 还可以帮助辅助优化编译。比如 Facebook 使用强化学习优化 LLVM 编译选项，性能提升了 15%。还可以自适应编译，即编译器根据运行时数据动态调整代码（如针对边缘设备的低功耗模式）。

随着 AI 的不断发展，编译技术也不断引入 AI 的优势，加速了编译技术的发展与革新。反之，编译技术的发展也给 AI 模型的部署带来了极大便利。如此相辅相成，我相信在未来编译技术一定会因为 AI 的发展而更上一层楼，而 AI 也会因为应用优秀的编译技术得以加速发展。未来，中国若能在这一领域抢占先机，或许能推动具有全球影响力的技术突破。

五、总结

综上所述，尽管现有开源编译器已经相当成熟，但中国仍然需要投入编译器技术研究和开发。这不仅是技术自主的需要，也是推动创新发展的机遇。同时，随着人工智能技术的发展，编译技术将在其中发挥越来越重要的作用。在重视安全风险的同时，我们应该积极推动技术创新，将编译与 AI 深度融合，为计算机科学和人工智能的进步做出贡献。

通过这次调研，我不仅对编译器技术有了更深入的了解，也更加明确了自己作为计算机专业学生的责任和使命。作为计算机科学与技术专业的学生，我们即将成为中国建设的中流砥柱。在包括编译技术在内的诸多高精尖领域，中国仍有广阔的提升空间需要我们探索。例如，华为在芯片和编译器研发中顶住外部压力取得的成就，表明自主创新的可行性，这对我们而言是一个值得学习的范例。为了更好地服务国家发展，我们应在夯实专业知识的基础上，主动思考如何将所学应用于实践。通过本次调研，我对编译器技术有了更深刻的理解，也更加明确了自身的责任与使命。我认为，国产编译器的发展不仅需要技术突破，还需构建完善的生态体系，这要求我们在学习中注重底层原理的掌握，并关注技术应用的实际场景。未来，我希望通过深入学习和参与相关项目，为国产编译器技术的发展贡献力量，同时为中国在全球技术竞争中赢得更多主动权尽一份责任！

参考文献

《中国基础软件发展白皮书》（中国信通院，2023）

LLVM 官方文档: <https://llvm.org/>

华为方舟编译器技术架构解析, 《软件学报》, 2022

"Supply Chain Attacks on Open Source Compilers", IEEE S&P, 2023

"MLIR: A Compiler Infrastructure for the End of Moore's Law", Google Research, 2021

Chris Lattner, Vikram Adve. "LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation"

Richard M. Stallman. "The GNU Compiler Collection"

中国计算机学会. "中国编译技术发展研究报告"

Alex Aiken. "Compiler Security and Optimization"

清华大学编译技术研究组. "人工智能编译优化研究进展"