Аннотация – в документ приводится подробный анализ уязвимости "LeftoverLocals" CVE-2023–4969, которая имеет значительные последствия для целостности приложений с графическим процессором, особенно для больших языковых моделях (LLM) и машинного обучения (ML), выполняемых на затронутых платформах с графическим процессором, включая платформы Apple, Qualcomm, AMD и Imagination.

Этот документ предоставляет ценную информацию для специалистов по кибербезопасности, команд DevOps, ИТ-специалистов и заинтересованных сторон в различных отраслях. Анализ призван углубить понимание проблем безопасности графических процессоров и помочь в разработке эффективных стратегий защиты конфиденциальных данных от аналогичных угроз в будущем.

I. Введение

Компания Trail of Bits раскрыла уязвимость под LeftoverLocals, которая позволяет восстанавливать данные из локальной памяти графического процессора, созданные другим процессом. Эта уязвимость затрагивает графические процессоры Apple, Qualcomm, AMD и Imagination и имеет значительные последствия для безопасности приложений на графических процессорах, особенно больших языковых моделях (LLM) и машинного обучения (ML), работающих на затронутых платформах.

Уязвимость позволяет злоумышленнику прослушивать интерактивный сеанс LLM другого пользователя несмотря на разграничения процессов и контейнеров., особенно в контексте LLMS и моделей ML, поскольку может привести к утечке конфиденциальных данных, участвующих в обучении этих моделей.

II. Уязвимая среда

Уязвимость LeftoverLocals может использоваться в различных средах, включая облачных провайдеров, мобильные приложения и даже при удалённых атаках.

- **Облачные провайдеры**: облачные провайдеры часто предлагают своим клиентам ресурсы графического процессора, которые используются совместно несколькими пользователями. В таких средах LeftoverLocals может быть использована при наличии доступа к тому же физическому графическому процессору, что и жертва. Это может позволить злоумышленнику восстановить данные из локальной памяти графического процессора, которые были созданы другим процессом, что приведёт к значительной утечке данных. Это особенно актуально для приложений, использующих LLM и ML для обработки данных.
- **Мобильные приложения**: Мобильные устройства, использующие уязвимые графические процессоры, также подвержены риску. Например, Apple признала, что такие устройства, как iPhone 12 и M2 MacBook Air, подвержены уязвимости LeftoverLocals..
- **Удалённые атаки**: LeftoverLocals потенциально можно использовать удалённо, когда злоумышленник скомпрометировал систему и получил возможность запуска пользовательского кода, либо в средах, где пользователи могут запускать пользовательские GPU вычислительных приложений.

III. LEFTOVERLOCALS В СРАВНЕНИИ С ДРУГИМИ УЯЗВИМОСТЯМИ

A. *leftoverlocals и другие GPU-уязвимости*

Уязвимость LeftoverLocals отличается от других уязвимостей GPU прежде всего методом утечки данных через локальную память GPU. В отличие от многих уязвимостей, которые используют определённые программные или аппаратные ошибки, LeftoverLocals основана на неспособности графических процессоров полностью изолировать память между процессами. Это позволяет злоумышленнику запускать вычислительное приложение на графическом процессоре для чтения данных, оставленных в локальной памяти графического процессора другим пользователем.

Другие же уязвимости графического процессора могут быть нацелены на различные аспекты архитектуры или программного обеспечения графического процессора, такие как переполнение буфера, или эксплойты на уровне драйверов. Эти уязвимости часто требуют выполнения определённых условий или зависят от сложного взаимодействия программного и аппаратного обеспечения.

Утечка данных может быть достаточно существенной для восстановления моделей или ответов, что представляет значительный риск для конфиденциальности обрабатываемой информации.

Факторы критичности уязвимости LeftoverLocals:

- **Широкое воздействие**: Уязвимость затрагивает широкий спектр графических процессоров крупных производителей, таких как AMD, Apple, Qualcomm и Imagination Technologies.

- **Утечка данных**: Например, на графическом процессоре AMD Radeon RX 7900 XT может произойти утечка около 5,5 МБ данных за один вызов графического процессора, что может составлять около 181 МБ для каждого LLM-запроса. Этого достаточно для восстановления отклика LLM с высокой точностью.
- **Простота использования**: уязвимостью можно воспользоваться, просто запустив приложение для вычислений на графическом процессоре для чтения данных, оставленных в локальной памяти графического процессора другим пользователем.
- **Проблемы с устранением уязвимости**: Устранение уязвимости может оказаться трудным для многих пользователей. Одним из предлагаемых решений является изменение исходного кода всех ядер GPU, использующих локальную память, для сохранения 0 в любых ячейках локальной памяти, которые использовались в ядре до его завершения. Однако это может повлиять на производительность.
- **Раскрытие конфиденциальных данных**: Уязвимость актуальна в контексте ИИ и машинного обучения, где конфиденциальные данные часто используются при обучении моделей.

B. *leftoverlocals и другие CPU-уязвимости*

Spectre и Meltdown являются уязвимостями ЦП, использующими атаки по "побочным каналам", которые включают извлечение информации из физической реализации компьютерных систем, а не программных ошибок. Spectre позволяет другим приложениям получать доступ к произвольным местоположениям в их памяти. Meltdown, с другой стороны, нарушает фундаментальную изоляцию между пользовательскими приложениями и операционной системой, позволяя приложению получать доступ ко всей системной памяти, включая память, выделенную для ядра.

Все три уязвимости серьёзны, поскольку могут привести к несанкционированному доступу к конфиденциальным данным. Однако они различаются по своему охвату и характеру данных, которые они могут раскрывать. LeftoverLocals в первую очередь влияет на приложения с графическим процессором и может привести к утечке данных из LLM и ML-моделей. Spectre и Meltdown, с другой стороны, потенциально могут раскрыть любые данные, обрабатываемые CPU, включая пароли, ключи шифрования и другую конфиденциальную информацию.

Потенциальные последствия уязвимостей:

- Они затрагивают почти все процессоры, выпущенные с 1995 года, что делает их влияние чрезвычайно распространённым.
- Они потенциально могут раскрыть любые данные, обрабатываемые центральным процессором, включая пароли, ключи шифрования и другую конфиденциальную информацию.
- Их трудно обнаружить, поскольку эксплуатация не оставляет никаких следов в традиционных файлах журналов.

C. *Сходство признаков уязвимостей*

LeftoverLocals имеет некоторое сходство с Spectre и Meltdown с точки зрения их последствий для безопасности:

- **Утечка данных**: как LeftoverLocals, так и Spectre / Meltdown допускают несанкционированный доступ к конфиденциальным данным. LeftoverLocals позволяет восстанавливать данные из локальной памяти графического процессора, в то время как Spectre и Meltdown используют спекулятивное выполнение ЦП для доступа к защищённой памяти.
- **Использование аппаратных возможностей**: Оба набора уязвимостей используют аппаратные возможности, предназначенные для оптимизации производительности.
- **Нарушение разграничения процессов**: оба механизмы обходят механизмы изоляции процесса для считывания данных на графических и центральных процессорах соответственно.
- **Влияние на нескольких поставщиков**: Обе уязвимости влияют на продукты нескольких поставщиков. LeftoverLocals влияет на графические процессоры Apple, Qualcomm, AMD и Imagination Technologies, в то время как Spectre и Meltdown влияют на процессоры Intel, AMD и ARM.
- **Смягчение последствий**: Устранение обеих уязвимостей является нетривиальным. LeftoverLocals может потребовать внесения изменений в код ядра GPU, в то время как Spectre и Meltdown потребовали сочетания обновлений микрокода, исправлений операционной системы и, в некоторых случаях, редизайна оборудования.
- **Скрытый характер атак**: Атаки, использующие эти уязвимости, трудно обнаружить, поскольку они не оставляют традиционных следов в файлах журналов, что затрудняет определение того, использовались ли они в реальных атаках.

IV. Требования к эксплуатации LeftoverLocals

A. *Общий доступ к графическому процессору*

Для использования LeftoverLocals требуется общий доступ к графическому процессору, что является обычным сценарием в многопользовательских средах, где несколько пользователей или приложений могут использовать одни и те же физические ресурсы графического процессора. Например, на платформах облачных вычислений, в общих центрах обработки данных или в любой системе, где GPU-ресурсы динамически распределяются между различными пользователями или задачами. В таких средах локальная память графического процессора не всегда очищается между различными исполнениями ядра или между использованием разными процессами.

B. *Модель "Listener-Writer"*

Модель Listener-Writer — это метод, используемый для эксплуатации уязвимости LeftoverLocals. Эти программы взаимодействуют с локальной памятью графического процессора, чтобы продемонстрировать уязвимость.

Writer служит для преднамеренного сохранения определённых canary-значений в локальной памяти графического процессора. Эти значения уникальны и идентифицируемы, они служат маркерами, которые могут быть обнаружены позже. Программа Writer не удаляет эти значения из локальной памяти после завершения своего выполнения.

Listener служит для чтения неинициализированной локальной памяти на графическом процессоре. Она сканирует локальную память в поисках значений canary, которые оставила программа записи. Если обнаруживает эти значения, это указывает на то, что локальная память не была должным образом очищена между выполнением разных программ.

C. *Доступ к устройствам*

Доступ к устройствам подразумевает определённый уровень доступа к ОС на целевом устройстве, чтобы воспользоваться уязвимостью. Этот доступ не обязательно должны быть root или администратор; это может быть любой уровень доступа, что позволяет злоумышленнику выполнить GPU-приложения.

В случае устройств Apple компания признала, что такие устройства, как iPhone 12 и M2 MacBook Air, подвержены уязвимости LeftoverLocals. Несмотря на то, что Apple выпустила исправления для своего новейшего оборудования, миллионы существующих устройств, использующих кремний Apple предыдущих поколений, остаются потенциально уязвимыми.

Qualcomm и AMD также подтвердили влияние уязвимости на свои графические процессоры и предприняли шаги по ее устранению. Qualcomm выпустила исправления для прошивки, а AMD имеет подробные планы по предоставлению дополнительных улучшений

V. Технологический процесс и PoC

A. *Модификация*

Первым шагом является изменение кода ядра графического процессора для чтения и записи в локальную память графического процессора, что позволяет создать PoC для прослушивания интерактивного сеанса LLM другого пользователя.

B. *Получение признаков LLM*

Получение признаков модели включает идентификацию конкретной используемой LLM путём наблюдения за шаблонами использования памяти графического процессора LLM. Разные LLM будут иметь разные схемы использования памяти, и, наблюдая за этими шаблонами, злоумышленник может определить, какая LLM используется. Информация может быть использована для адаптации атаки к конкретному LLM, повышая шансы на успешное использование уязвимости.

C. *Прослушивание выходных данных LLM*

После получения признаков модели злоумышленник может начать прослушивание выходных данных LLM. Это включает в себя повторный запуск GPU-ядра, которое считывает данные из неинициализированной локальной памяти на графическом процессоре. Далее сканируется локальная память в поисках определённых значений, оставленных LLM. Их обнаружение указывает на то, что локальная память не была должным образом очищена между выполнением различных программ. Это позволяет восстановить данные из вычислений LLM.

D. *PoC*

PoC разработан с использованием фреймворка-OpenCL, фреймворка для выполнения на разнородных платформах для демонстрации ключевых особенностей:

- **Получение признаков модели**: PoC включает в себя идентификацию конкретной используемой LLM путём наблюдения за шаблонами использования GPU-памяти. Разные LLM имеют разные схемы использования памяти, по которым можно определить, какой LLM используется.
- **Прослушивание выходных данных LLM**: PoC включает повторный запуск GPU-ядра, которое считывает данные из неинициализированной локальной памяти на графическом процессоре. Злоумышленник сканирует локальную память в поисках определённых значений, оставленных LLM. Если эти значения обнаружены, это указывает на то, что локальная память не была должным образом очищена между выполнением различных программ, что позволяет злоумышленнику восстановить данные из вычислений LLM.
- **Утечка данных**: обнаружено что LeftoverLocals приводит к утечке ~5,5 МБ за каждый GPU-вызов на AMD Radeon RX 7900 XT, что при запуске модели 7B составляет около 181 МБ за каждый запрос LLM. Этой информации достаточно для восстановления ответа LLM с высокой точностью.
- **Обход механизмов изоляции процессов и контейнеров**: PoC демонстрирует, что злоумышленник может прослушивать интерактивный сеанс LLM другого пользователя в обход изоляции процесса или контейнера. Это показывает, что уязвимость может быть использована в многопользовательских средах, таких как платформы облачных вычислений, где несколько пользователей используют один и тот же физический графический процессор.
- **Доступ к устройствам**: PoC требует, чтобы злоумышленник имел доступ к целевому устройству. Это может быть любой уровень доступа, позволяющий злоумышленнику выполнять свои собственные вычислительные приложения на графическом процессоре.