

ОТЗЫВ

официального оппонента на диссертацию

Иваничкиной Людмилы Владимировны

«Математические модели надёжности и методы её повышения

в современных распределённых отказоустойчивых системах хранения данных»,

представленную на соискание учёной степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Актуальность избранной темы. Надёжное хранение накопленных данных – одна из перманентно актуальных проблем человечества. В настоящее время в связи с развитием технологий обработки «больших данных» и облачных технологий. Наиболее популярные по соотношению скорости, размера и стоимости дисковые накопители имеют весьма ограниченную (относительно типичных объёмов хранилищ) ёмкость и подвержены нескольким классам отказов, поэтому надёжное хранение данных можно обеспечить, лишь объединив ресурсы множества дисков и организовав автоматическое восстановление после дисковых отказов за счёт хранения избыточной информации. Системы хранения данных (СХД) прошли долгую эволюцию от локальных *RAID*-массивов до современных хранилищ данных облачных платформ, объединяющих много тысяч дисков. Однако теоретические основы обеспечения надёжности хранения данных в таких системах разработаны явно неадекватно современным и перспективным СХД. Так, в настоящее время доминирующей имитационной моделью является Марковская модель надёжности, хотя в реальных СХД она применима лишь качественно, поскольку описывает набор состояний единственного фрагмента данных. В рассматриваемой работе автором сделана попытка выйти за рамки классической Марковской модели и разработать класс математических моделей, адекватных свойствам современных СХД.

Содержание работы

Работа из 114 страниц содержит введение, четыре главы, заключение и список литературы.

Во введении обосновывается актуальность темы исследования, научная и практическая ценность работы, научная новизна, представлены выносимые на защиту результаты и кратко изложена структура диссертации.

В первой главе определены и детально проанализированы основные требования к современным системам хранения данных, разобраны преимущества и недостатки существующих подходов и решений.

Во второй главе изучены математические модели надёжности систем хранения данных. Описана классическая Марковская модель надёжности хранения данных, приведено аналитическое решение задачи о среднем времени наработки до отказа. Рассмотрен оригинальный метод решения задачи нахождения среднего времени наработки до отказа в приближении малости интенсивности отказов по сравнению со скоростью восстановления для произвольного графа состояний системы.

Рассмотрены причины того, что результаты, полученные на основе Марковской модели, лишь качественно применимы для описания реальной СХД. Так, Марковская модель описывает независимую эволюцию (с акцентом на деградации) единственного фрагмента данных, а в реальной СХД существует большое число фрагментов данных, которые в иерархически организованной структуре единиц хранения и управления, что приводит к сильным зависимостям свойств фрагментов друг от друга. В частности, отказ диска приводит к утрате всех хранившихся на нём блоков данных, а восстановление утраченных блоков происходит не независимо (абсолютно параллельно), а в основном последовательно, причём со скоростью, ограниченной суммарной производительностью некоторого подмножества дисков СХД.

Учёт упомянутых выше факторов одновременно с отказом от Марковской модели, описывающей эволюцию отдельного фрагмента данных, позволил автору предложить оригинальную математическую модель, описывающую надёжность СХД в целом. Для проверки предсказаний разработанной математической модели автором рассмотрены две имитационные модели, имитирующие реальные СХД с различной детальностью.

В третьей главе рассмотрены факторы, влияющие на надёжность хранения данных в реальных СХД, такие как политики размещения дисковых блоков (группы размещения), наличие скрытых повреждений дисковых блоков, а также учёт особенностей аппаратной инфраструктуры (области отказов). Математическая модель надёжности дополнена с целью адекватного учёта групп размещения и скрытых повреждений. Доказана важная для практической реализации СХД теорема, дающая нижнюю оценку вероятности потери данных в единицу времени при наличии скрытых повреждений.

В четвёртой главе приведено описание внедрения результатов исследования надёжности СХД, которые нашли применение при создании кластерного хранилища данных, вошедшего в состав программных комплексов *Acronis Storage*, *Virtuozzo Storage* и *P*-платформа. Отметим, что дан хороший обзор направлений развития моделей и значимости этого развития, причём все направления непосредственно связаны с трендами в области СХД.

В заключении приведены основные результаты работы.

Новизна исследований и полученных результатов

В качестве научных результатов отметим следующие.

- Предложена оригинальная математическая модель надёжности СХД с разбиением данных на блоки, адекватная реально существующим современным системам хранения данных. Модель хорошо интерпретируется
- Разработан комплекс имитационных моделей надёжности СХД, отражающий реальную архитектуру современных СХД.
- Детально исследовано влияние различных факторов на надёжность и масштабируемость СХД. Теоретически изучено и проверено с помощью имитационных моделей влияние различных политик размещения дисковых блоков на надёжность хранилища (в зарубежных публикациях по теме аналогов не найдено). Получены количественные оценки влияния скрытых повреждений на надёжность СХД. Этот результат особенно ценен с точки зрения практического применения предложенных имитационных моделей.
- На базе полученных теоретических результатов построена распределённая система хранения данных, гарантирующая высокую надёжность за счёт использования различных схем обеспечения избыточности.

Теоретическая значимость и практическая полезность

Теоретической значимостью обладает новый класс математических моделей СХД, более адекватных реальным СХД по сравнению с Марковскими моделями, а также исследование характеристик и границ применимости моделей. Ещё раз отметим хорошо описанные возможности дальнейшего развития моделей по нескольким направлениям.

Результаты теоретических исследований применены, например, при создании кластерного хранилища «*Acronis Storage*» компании *Acronis*, которое является ключевым участником мирового рынка, что отлично иллюстрирует практическую полезность.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

Выводы, сформулированные в результате проведённого исследования, обоснованы. Для теоретических результатов приведены достаточно полные доказательства. Приведённые автором рекомендации проверены апробацией и внедрением.

Достоверность и новизна исследования, полученных результатов, выводов и рекомендаций, сформулированных в диссертации

Достоверность научных исследований и результаты работ подтверждаются корректным анализом теоретических построений и проверкой полученных результатов на имитационных моделях надёжности СХД, отражающих реальную архитектуру и особенности реализации подобных систем.

Недостатки работы

- Не хватает выделенного глоссария, особенно в связи со слегка нестандартным использованием некоторых терминов. Отмечу собственно надёжность (в разных значениях), отказ (в двух значениях), область отказа, деградацию, скраббинг, чанк, страйп и др.
- Приоритизация (восстановления, скраббинга, ...) как одна из особенностей модели должна быть описана более явно.
- В четвёртой главе не хватает хорошей иллюстрации компромисса между стоимостью, объёмом, производительностью, масштабируемостью и другими параметрами в «численном» виде с графиками и диаграммами. Это бы заметно поддержало сформулированные рекомендации.
- Напрашивается хотя бы базовый анализ интересных (и достаточно простых) предложений по расширению в разделе 4.4.2, их привязка к модели и учёт влияния на модель. Под привязкой к модели имеется в виду усложнение схемы хранения путём явного задания параметров терархии областей отказов, управления метаданными и процедур восстановления.

Заключение

Диссертация Л.В. Иваничкиной на соискание учёной степени кандидата технических наук является научно-квалификационной работой, в которой содержится решение задачи повышения надёжности систем хранения данных, имеющей существенное значение для развития информационной инфраструктуры.

Полученные автором результаты достоверны, выводы и заключения обоснованы. Основные научные результаты работы отражены в восьми публикациях, среди которых рекомендованные ВАК, индексируемые системами *Web Of Science* и *SCOPUS*, два свидетельства о регистрации программы и заявка на патент. Автореферат достаточно полно и верно отражает содержание и основные результаты работы.

На основании вышеизложенного можно заключить, что диссертация Л.В. Иваничкиной «Математические модели надёжности и методы её повышения в современных распределённых отказоустойчивых системах хранения данных» является законченной научно-квалификационной работой и удовлетворяет требованиям «Положения о порядке присуждения учёных степеней», утверждённого постановлением Правительства Российской Федерации от 24.09.2013 г. №842, предъявляемым на соискание учёной степени кандидата наук, а её автор заслуживает присуждения учёной степени кандидата технических наук по специальности

05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Официальный оппонент,

старший научный сотрудник международной научно учебной лаборатории интеллектуальных систем и структурного анализа факультета компьютерных наук национального исследовательского университета «Высшая школа экономики», к.т.н., доцент

Незнанов Алексей Андреевич

06 04. 2018 .