

К ОДНОКРИСТАЛЬНЫМ КОМПЬЮТЕРАМ-УСКОРИТЕЛЯМ С МАСШТАБИРУЕМЫМ МНОГОЯДЕРНЫМ ПАРАЛЛЕЛИЗМОМ ДЛЯ СОЗДАНИЯ ЦИФРОВЫХ ДВОЙНИКОВ БОЛЬШИХ СИСТЕМ (НА ПРИМЕРЕ АЭС)

Затуливетер Ю.С., Фищенко Е.А.

Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия

zvt@ipu.ru, elena.fish@mail.ru

Аннотация. Для критически важных классов распределённых техногенных систем, в том числе из энергетических сфер АЭС, обозначены направления и пути развития больших компьютерно-сетевых вычислительных сред с облакоцентричными инфраструктурами и методов реализации в них цифровых двойников с применением отечественной элементной базы гибридных компьютерных архитектур.

Ключевые слова: глобальная компьютерная среда, цифровые двойники, АЭС, клиент-серверные сетевые архитектуры, облачные и пост-облачные системы, виртуализация, импортонезависимость, однокристалльный многоядерный компьютер-ускоритель.

Введение

Перспективы становления и развития концепции, методов и технологий цифровых двойников (ЦД) диктуются глубинными процессами роста глобальной компьютерной среды (ГКС) и всеохватывающим расширением сфер её влияния. В течение трёх предшествующих десятилетий посредством ГКС де-факто без продуманной стратегии осуществляется погружение социотехносферы в глобально сильносвязное информационное пространство, с исторически беспрецедентной метрикой, выражаемой формулой «Всё-влияет-на-всё-и-сразу» [1]. При этом цифровая трансформация сильносвязных, всё более крупных, динамичных и неустойчивых (потому трудно управляемых) распределённых социотехногенных систем в глобально сильносвязную киберсоциофизическую метасистему [2] до сих пор осуществляется стихийно и фрагментировано.

Недостаточные масштабы комплексно-многофакторного предупреждающего моделирования деструктивных последствий ведут к глобализации растущих издержек от побочных эффектов. Свидетельство – разнообразные и глобально прогрессирующие проявления кризисов, разрушающих традиционные («доцифровые») модели и институты управления устойчивым развитием.

Возможные причины техногенного характера – во-первых, изначально отсутствие в сколь угодно больших изначально разнородных сетевых ресурсах высокоорганизованных распределённых вычислительных сред с универсальными системными свойствами структурной и вычислительной масштабируемости. Во-вторых, – отсутствие сопутствующих универсальных технологий для сквозного программирования масштабируемых и увязанных между собой разнопрофильных ЦД в сколь угодно больших сетях. Такие среды и технологии должны отвечать растущим требованиям и кибербезопасности, и «бесшовной» программируемости, и системно-целостного моделирования больших взаимосвязанных и всё более сложных распределённых социотехногенных систем [3, 4].

В результате спонтанно-взрывного, потому системно несогласованного и несбалансированного, расширения ГКС и сфер её влияния, кусочно-фрагментарное осуществление массовой цифровой трансформации сопровождается побочными эффектами и сопутствующими издержками в виде роста киберугроз, снижения устойчивости и эффективности функционирования/развития больших распределённых систем [3], включая техногенные системы индустриальной направленности [4].

В условиях глобальной информационной сильносвязности [1] это прямо или косвенно относится не только к отдельным видам таких систем, но и ко всему быстро растущему разнообразию систем, охватываемому концептом «Индустрия 4.0», который в настоящее время распространяет свои принципы и в весьма консервативные сферы атомной энергетики [4 - 7].

В работе с общесистемных позиций анализируются архитектурно-системотехнические аспекты развития ГКС. В ориентации на критические классы техногенных систем обозначаются направления развития сетевых вычислительных сред и методов реализации в них ЦД с применением гибридных компьютерных систем высокой производительности на основе отечественной элементной базы (ЭБ).

Рассмотрены пути такого развития для облачных сетевых вычислительных сред. В среднесрочной перспективе – это импортонезависимое продвижение мирового тренда развития пост-облачных систем. В целях расширения сфер применения облачных систем они осуществляют их частичную децентрализацию путём добавления периферийных сетевых слоёв. Привнесение в серверные и

периферийные ресурсы однокристалльных компьютеров-ускорителей с массовым многоядерным параллелизмом открывает новые направления пост-облачного развития.

1. О перспективах развития концепции цифровых двойников

Из сказанного можно заключить, что сверхзадача отечественной линии развития концепции ЦД, а также суверенных технологий их воплощения состоит в том, чтобы динамично, с приоритетами на критических направлениях развития техносферы, группировать и направлять неограниченно растущий системообразующий и вычислительный потенциал глобальной цифры, концентрируемый в ГКС, на обеспечение безопасности процессов цифровой трансформации растущего разнообразия социотехногенных систем в глобально сильносвязную и системно сбалансированную киберсоциофизическую метасистему [2]. В беспрецедентных условиях глобальной информационной сильносвязности необходимо должным образом обеспечивать соответствие растущим требованиям к кибербезопасности и управлению их устойчивым функционированием/развитием [1, 3, 4].

Многие критически важные отрасли промышленности, включая сферы производства энергии, в том числе атомная энергетика, опираются на весьма дорогостоящие и потенциально небезопасные физические объекты своей инфраструктуры. В настоящее время это – всё более крупные и структурно-сложные распределённые системы, управление которыми осуществляется с применением цифровых компьютерно-сетевых средств и технологий [4, 8]. В рамках концепта «Индустрия 4.0» такие системы на всех этапах своего жизненного цикла (проектирование, функционирование, эксплуатация, модернизация, утилизация) требуют комплексно-многофакторного, всё более «тяжёлого» [4] обеспечения необходимых уровней безопасности и эффективности.

В качестве иллюстрации к сказанному на рис. 1 на примере АСУ ТП АЭС показаны многофакторные взаимосвязи между этапами жизненного цикла больших систем.



Рис. 1. Жизненный цикл АСУ ТП АЭС [6] (http://images.myshared.ru/6/691812/slide_2.jpg)

Для достижения требуемых уровней безопасности и эффективности в условиях растущей с опережением сложности всё более крупных мультифункциональных систем применяются и развиваются методы моделирования в русле концепции и технологий ЦД. Концептуальный характер моделирования посредством технологий ЦД придаёт функциональную направленность методам и средствам виртуализации сетевых ресурсов [4 - 15], как воплощаемых в «металле» систем, так и процессам их цифровой трансформации в общем случае в крупные киберсоциофизические системы.

Технологии моделирования сложных систем с применением ЦД предполагают создание/развитие в разнородных вычислительных средах растущего количества разнопрофильных, масштабируемых и взаимоувязанных цифровых моделей физических узлов, агрегатов и систем, которые способны совместно с реальными объектами обрабатывать в реальном времени реальные данные, дублируя и контролируя в цифровом пространстве их функционирование.

В концепции ЦД компьютерные модели (включая динамические) должны с требуемой степенью приближения не только в автономно-модельном и технологическом масштабах времени отображать,

но и прогнозировать в упреждающих реальное время режимах функционирование растущего количества взаимосвязанных объектов в целях обеспечения безопасности и эффективности распределённых систем. Это необходимо для контроля безопасности процессов их разработки, изготовления, а также управления функционированием как в штатных, так и в нештатных режимах.

Актуальность упорядочения глобальных процессов цифровой трансформации в промышленных сферах, осуществляемого путём системно согласованного и сбалансированного увеличения сетевых масштабов и глубины приближения ЦД к реалиям, растёт с использованием оборудования высокой сложности, а также при работе с потенциально опасными объектами, где доминируют сложные условия эксплуатации и агрессивные среды, ограничивающие возможности установки измерительного оборудования. В таких системах потребность развития методов и компьютерно-сетевых средств цифрового моделирования возрастает многократно. Это в полной мере справедливо для задач, с которыми сталкиваются и специалисты атомной энергетической отрасли [4 - 15].

Стимулами к созданию ЦД являются [8]:

- количественный рост систем, сопровождаемый увеличением структурной/функциональной сложности создаваемых инженерных объектов (требует новой парадигмы проектирования, позволяющей преодолеть барьеры между дезинтегрированными средствами разработки, осуществить интеграцию среди групп разработчиков, что необходимо для достижения растущих уровней безопасности и эффективности);
- дополнительные возможности повышения конкурентных преимуществ промышленных изделий, достигаемых за счёт снижения затрат и ускорения выпуска продукции, создания более надёжных и менее материалоёмких конструкций;
- понимание того, что дополнительные усилия по созданию ЦД на стадиях разработки ведут к общему снижению расходов не только на стадиях проектирования, но и на всём протяжении жизненного цикла изделий, включая эксплуатацию, мониторинг, модернизацию и утилизацию.

Широкое внедрение ЦД, с растущими масштабами их системно-функционального согласования, позволяет в ходе виртуальных испытаний вариантов решений по разрабатываемым узлам, агрегатам и системам выносить верификацию ответственных решений на начальные этапы проектирования. Упреждающее внесение необходимых изменений сокращает затраты средств и времени. При этом создание двойников только на этапах проектирования недостаточно [4 - 7]. В ходе эксплуатации во многих случаях требуется построение взаимоувязанных, непрерывно функционирующих в реальном и упреждающем времени моделей различных этапов жизненного цикла изделий.

Распределённые системы реализуются в вычислительных средах компьютерных сетей, которые изначально составлены из крайне разнородных, потому трудно совместимых аппаратных, программных и информационных ресурсов. В силу комбинаторной сложности многовариантных задач системно-функциональной интеграции разнородных сетевых ресурсов с увеличением размеров распределённых систем и, соответственно, размеров функционально ориентированных сетевых вычислительных сред системотехнические сложности интеграции носят фундаментальный (вне технологический) характер и растут со сверхлинейным опережением [2, 3].

В условиях разнородности сетевых ресурсов для снижения затрат на преодоление комбинаторных барьеров сложности системно функциональной интеграции есть два пути. Первый (ближняя и среднесрочная перспектива) – неизбежное ограничение размеров сетевой среды, второй – устранение причин непрерывного воспроизводства разнородности в сколь угодно больших сетях. Первый путь – это текущий мейнстрим развития облакоцентричных систем. Результаты исследований перспектив движения по альтернативному пути представлены в [1 - 3, 16, 17, 19, 20] и других публикациях авторов данной работы.

С увеличением размеров распределённых систем при их структурных трансформациях проблемы системно-функциональной интеграции сетевых ресурсов требуют особых решений. Для преодоления барьеров комбинаторной сложности при создании и широком использовании ЦД большие разнородные вычислительные среды, воплощаемые в сетях, должны обладать регулярными свойствами системной и вычислительной масштабируемости, достигаемые посредством виртуализации компьютерных и сетевых ресурсов [7, 11, 13]. Эти дополнительно привносимые и весьма непростые в реализации системные требования необходимы для снижения затрат на системно-функциональную интеграцию разнородных ресурсов до приемлемых уровней.

В стратегии своей эволюции ЦД должны обеспечивать формирование в инфраструктуре ГКС единого (сквозного) информационно-алгоритмического пространства [2, 3], которое позволит с минимальными издержками системно-целостным образом связывать физические реалии и отражающие их виртуальные образы параллельно формируемого цифрового мира.

2. Эволюция глобальной цифры: стадии развития сетевых инфраструктур

Рассмотрение стадий развития ГКС ведётся в привязке к широко распространённой их маркетинговой маркировке в виде мемов терминологического ряда Web 1.0, Web 2.0, Web 3.0, идентифицирующим три стадии эволюции всемирной паутины WWW (рис. 2). По времени эти стадии охватывают (с нахлёстом) три десятилетия, начиная с 90-х, в которых появился и глобально сформировался беспрецедентный технологический феномен WWW.



Рис. 2. Маркетинговая версия эволюции Интернета (<https://cryptotrek.ru/wp-content/uploads/2022/11/webs.jpg>)

Каждое из трёх предшествующих десятилетий расширения сфер влияния экспоненциально растущей ГКС оказало своё глобальное воздействие на социотехносферу. Это влияние отражено на рис. 2 тремя стадиями роста принципиально новых информационных услуг массового потребления.

Многие авторы идентифицируют эти этапы различающимися, но во многом схожими, наборами доминирующих характеристик оказываемых услуг (<https://habr.com/ru/articles/653533/16>). Они перефразируют или расширяют трактовки, изначально предложенные авторами маркетинговых терминов, с которыми стартовали кампании продвижения услуг (на рис. 2 обозначены значками) на рынки. При этом растущее внимание на 2-3 стадиях уделяется влиянию на сферы массового потребления быстрорастущего разнообразия принципиально новых – информационно-алгоритмических – услуг. Это разнообразие сопутствует (как следствие) системотехническому прогрессу ГКС, который обеспечивает неизменный рост спроса на расширение сфер её применения, что выражается в устойчивом прогрессе рынков глобально доступных цифровых услуг.

В основной своей части такая аргументация относится к сферам маркетингии наступающего цифрового мира (рис. 2). В этих сферах субъективные оценки и выводы фокусируются на растущих как грибы после дождя проявлениях внешних факторов рыночного потребления новых видов услуг. Понятно, что такие исследования не предполагают прямое использование научно-технической доказательной базы, с аргументацией, идентифицирующей и раскрывающей фундаментальные закономерности определяющие траектории глубинных процессов, лежащих в основе развития ГКС.

Очевидно, однобокий взгляд со стороны внешних сфер массового потребления не раскрывает сути глубинных процессов системотехнического развития глобальных сетей, поэтому не может дать полной картины для понимания движущих сил и этапов развития инфраструктуры и функционала собственно ГКС. Чтобы восполнить это упущение, мы делаем попытку формирования встречного общесистемного взгляда «из глубины» ГКС путём системотехнического анализа стадий её развития.

В настоящее время системотехническое развитие ГКС продолжается в рамках тренда расширения системообразующих возможностей сетевых архитектур – «Client-Server» (далее «Клиент-Сервер»). В таких архитектурах функционал переработки потоков клиентских запросов концентрируется в централизованных вычислительных ресурсах сетевых узлов на серверной стороне распределённых систем (с априори ограничиваемым количеством серверов). Число клиентских сетевых узлов может легко наращиваться путём их санкционированного присоединения посредством общедоступных или специальных средств отдалённого сетевого доступа.

В глобальных сетях клиент-серверные архитектуры прошли разные фазы своего развития. С архитектур этого типа началось массовое освоение глобальных сетей путём формирования всемирной паутины WWW. Посредством сетевой гипертекстовой модели пространство WWW предоставило массовый доступ со стороны клиентских браузеров к текстово-мультимедийному контенту, глобально распределённому по Web-серверам сетевых узлов. Эта модель легла в основу простейшей в реализации потребительской глобально-сетевой инфраструктуры, посредством которой был сделан первый шаг в формировании массовой сферы общедоступных информационных услуг по производству-потреблению распределённой информации. Количественный рост потоков/объёмов гипертекстовой информации сформировал беспрецедентный феномен общедоступности информации

с глобальной информационной сильносвязностью [1]. Вместе с тем практика освоения гипертекстового Web-пространства показала, что в условиях информационной сильносвязности для устойчивого функционирования и развития необходима более продуктивная модель работы с глобально распределённой информацией, в которой переработка заметной её части должна осуществляться в составе ГКС посредством алгоритмов с использованием быстро прогрессирующих компьютерных средств разных классов.

Такая модель была реализована в середине 00-х в рамках клиент-серверных сетевых архитектур посредством первых (<https://practicum.yandex.ru/blog/oblachnye-tehnologii/>) коммерчески значимых облачных систем. К тому времени в состав ГКС включаются быстро прогрессирующие в количестве и функционале общедоступные мобильные компьютерные устройства беспроводной коммуникации. Введение их в циклы массового производства-потребления информационно-алгоритмических услуг резко ускорило рост разнообразия сфер применения ГКС и масштабов её влияния.

Далее по отдельности рассмотрим стадии развития клиент-серверных сетевых инфраструктур ГКС с момента её появления.

2.1. Формирование и развитие пространства WWW

С начала 90-х годов становление и развитие ГКС (в части предоставления массовых информационных услуг) формировалось в парадигме глобально-сетевое гипертекстовое пространство WWW (<https://webfoundation.org/about/vision/history-of-the-web/>) в сетевой архитектуре «Клиент-Сервер» с гипертекстовой инфраструктурой <Web-сервер – Браузер>. Доминирующая (в 90-х) глобализация информационного пространства в рамках гипертекстовой парадигмы была изначально специализирована на неалгоритмическое восприятие/обработку информации человеком.

Главный парадокс и глобальный системный дисбаланс быстро растущих миллионов сетевых узлов ГКС, все узлы которой – это универсально программируемые компьютеры различных классов, в том, что из-за отсутствия алгоритмической универсальности гипертекста совокупный функционал ГКС, распределённый по её универсальным компьютерам, в применении к обработке распределённой информации оставался в большей своей части невостребованным. Образно говоря, универсальные системообразующие возможности ГКС в целом были «заморожены» принципиально ограниченным функционалом глобально распределённого «информационного калькулятора» [16].

В режимах общедоступного «калькулятора» ГКС стала активнейшим образом использоваться различными сообществами, в целях наполнения и освоения внезапно открывшегося глобально сильносвязного мирового информационного пространства WWW. В таком «калькуляторе» гиперссылки на клиентских страницах браузеров в режиме ввода данных на обработку заняли место клавиатуры привычных калькуляторов. При этом обработка гиперссылок, пересылаемых по сетям на Web-серверы, вместо функционала аппаратно встроенных в обычные калькуляторы арифметических операций сводится к простейшей в исполнении обратной пересылке на клиентские браузеры запрошенных страниц контента с другими гиперссылками.

В связи с вовлечением в это пространство массовых пользователей большой бизнес увидел в WWW практически неограниченные возможности сетевой глобализации универсальной рыночной площадки. Однако, растущий в 90-х в темпах роста числа пользователей WWW, энтузиазм инвесторов (разогретый ожиданиями быстрой гипертекстовой глобализации мирового экономического пространства) к 2000 году обернулся глобальным финансовым кризисом «доткомов» с потерей инвесторами более \$5трлн (<https://internetboss.ru/krisis-dotkomov/>).

Глубинные причины этого кризиса следует искать в системотехнических недрах ГКС. Отметим некоторые аспекты влияния системно несбалансированного роста ГКС.

Опережающий рост гипертекстовой информации, ориентированной на восприятие человеком, уже к концу 90-х на порядки превысил совокупные возможности сообществ Homo Sapiens в части субъектно-смысловой переработки совокупной информации в целях управления устойчивым развитием социотехносферы. Web-страницы, рассчитанные на трудно формализуемое восприятие человеком, в чрезмерных количествах накапливались на Web-серверах. Они оставались недоступным для компьютерной переработки в целях устойчивого развития и де-факто превращались в свалки информационных отходов. Понятно, что при таком накоплении глобально распределённой информации её системообразующий потенциал обесценивается. Для бескризисной глобализации сетевой цифровизации бизнес-процессов нужны модели массового предоставления информации в пакете с полновесными алгоритмическими услугами по её переработке.

К середине 00-х экспоненциально растущие в WWW потоки и объёмы такой информации, превысили критические уровни. Первые коммерчески значимые шаги в части алгоритмизации

обработки глобально распределённой информации сделали облачные системы. Такие системы стали принципиальным развитием глобально сетевых клиент-серверных архитектур. Они позволили вывести на мировой рынок комплексные информационно-алгоритмические услуги в одном пакете <информация+алгоритмика>, в которых алгоритмическая составляющая поддерживается растущей вычислительной мощностью универсально программируемых компьютеров серверной стороны.

2.2. Облакоцентричные сетевые архитектуры распределённых систем

Проблематика глобальной цифровой трансформации техногенных систем в киберфизические системы, функционирующие с требуемой безопасностью и эффективностью, представленная в концепте Индустрия 4.0, предполагает неограниченное использование растущего потенциала ГКС для реализации методов и технологий системно целостной виртуализации сложных распределённых систем промышленной направленности посредством применения методов и технологий ЦД. Очевидно, как следует из сказанного выше, что в гипертекстовой сетевой инфраструктуре глобального информационного калькулятора обозначенные проблемы цифровизации нерешаемы.

Наращивание масштабов распределённых информационных и вычислительных систем стало возможным на основе технологий интеграции сетевых ресурсов в рамках концепции облачных вычислений, которые основаны на централизованных клиент-серверных сетевых архитектурах.

Сетевые облачные системы открыли новые этапы глобализации универсально программируемой алгоритмической переработки распределённой информации. В качестве серверов используются мощные универсальные компьютеры с многоядерными архитектурами, дополненные сетевыми средствами связи между собой и внешним миром, а также системами долговременного хранения больших и сверхбольших объёмов данных. Отметим, глобализация алгоритмики – один из ключевых трендов развития системообразующих качеств ГКС на стадии Web 2.0 (рис. 2)

Термин Cloud Computing возник (<https://compress.ru/article.aspx?id=20755>) в 2001 году. В 2006 г. появились (<https://practicum.yandex.ru/blog/oblachnye-tehnologii/>) коммерческие облачные платформы Zimki и Amazon Web Services. Предоставив через Интернет в режиме «тонкого клиента» доступ к сервисам для вычислений и работы с данными, они впервые открыли общедоступный сетевой рынок информационно-алгоритмических услуг. В 2008 г Google запустил App Engine. В начале 10-х появились платформы Microsoft Azure и Google Cloud, с которых началось массовое освоение облачных технологий в разных сферах бизнеса (<https://practicum.yandex.ru/blog/oblachnye-tehnologii/>).

Облачные архитектуры (рис. 3а) позволяют формировать в сетевом окружении своего серверного центра большие развивающиеся экосистемы производства-потребления инфраструктурных (I-as-a-Service), платформенных (P-as-a-Service) и программных информационно-алгоритмических (S-as-a-Service) услуг (<https://www.corpsoft24.ru/about/blog/iaas-saas-paas-s-podrobnostyami-i-primerami/>). При этом пространство предоставляемых услуг каждой облачной системы остаётся замкнутым в своей ограниченной системно-функциональной направленности.

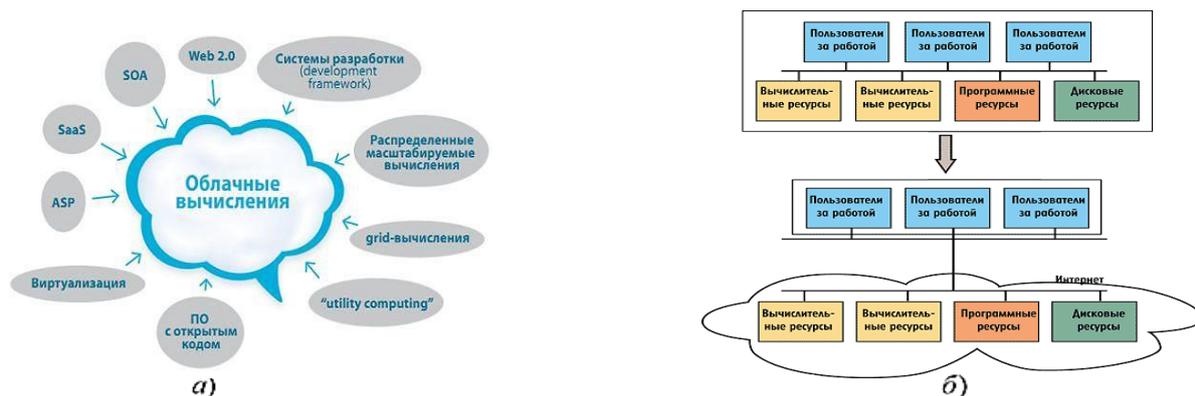


Рис. 3. Облачные вычисления: а) интеграция многих компьютерно-сетевых технологий (<https://sonikelf.ru/oblachnye-tehnologii-dlya-zemnyx-polzovatelej/>), б) виртуализация серверных ресурсов в целях масштабирования и обеспечения массового клиентского доступа к ним через Интернет (<https://compress.ru/article.aspx?id=20755>)

На рис. 3а показана общая схема облачных платформ с архитектурами «Клиент-Сервер», которые воплощают принцип асимметричной поляризации сетевых узлов таких систем путём пространственного разделения на серверную часть и клиентскую. В серверной части централизованным образом концентрируются компьютерные средства для разнообразной

алгоритмической обработки больших потоков клиентских запросов, а также для накопления всех объёмов значимых данных, циркулирующих в своих экосистемах. Такая централизация наряду с достоинствами имеет и принципиальные недостатки – риски утраты или компрометации больших объёмов данных, растущие нагрузки на каналы связи при увеличении числа клиентских узлов.

Клиентские узлы имеют санкционированный сетевой доступ к серверам и могут наращиваться в больших количествах. Алгоритмический функционал отдалённого доступа клиентских мест к серверному центру сужен до разнопрофильных интерфейсов, которые специализированы на предоставлении информационно-алгоритмических услуг в режимах «тонкого клиента».

Со второй половины 00-х прогресс архитектур облачных систем, с высокой рентабельностью предоставляющих массовые комплексные услуги <информация+алгоритмика>, до сих пор остаётся ведущим индустриальным трендом в развитии компьютерно-сетевых технологий, с помощью которых осуществляется очаговая цифровизация социотехносферы посредством формирования облачных экосистем различной направленности.

На рис. 36 показано, что системно-функциональная интеграция сетевых серверных ресурсов в разных облачных системах достигается применением методов и средств их виртуализации. Посредством виртуализации обеспечивается достижение системных свойств структурно-функциональной и вычислительной масштабируемости, которые необходимы для формирования, функционирования и развития своего агрегированного в каждой облачной экосистеме информационно-алгоритмического пространства обработки информации.

Важно отметить, что в ходе виртуализации разнородных сетевых ресурсов облачных систем проблемы их системно-функциональной интеграции замыкаются в априори ограниченном пространстве сетевых узлов серверной стороны. Преодоление комбинаторных барьеров сложности при этом осуществляется за счёт ограничения количества серверных узлов до приемлемых по затратам уровней и системотехнических решений по интеграции ресурсов. Эти уровни обеспечиваются индустриальными стандартами и поддерживающими эти стандарты технологиями виртуализации серверных ресурсов. Благодаря всегда ограниченному числу сетевых узлов серверных ресурсов, в облачных системах комбинаторный характер сложности проблем интеграции разнородных ресурсов перестаёт зависеть от в потенциале неограниченно растущего количества клиентских узлов. В этом состоит то главное фундаментальное качество, которое составляет основу облачного мейнстрима, который до сих пор обеспечивает наращивание размеров и масштабов применения больших распределённых систем. Однако, как показано в [17], в развитии облачных инфраструктур имеются принципиальные пределы роста размеров и масштабов применения.

Одним из главных ограничений дальнейшего развития облачных архитектур, предоставляющих пакеты услуг <информация+алгоритмика> стала чрезмерная централизация серверной обработки распределённой информации. Она показала свою эффективность в больших распределённых системах и большими потоками независимых запросов к серверам, в которых массовая клиентская часть отделима от серверной посредством функционирования в режимах «тонкого клиента». Такой режим предписывает полномасштабный перенос алгоритмического функционала в серверную часть.

Однако, такие облачные системы не отвечают требованиям киберфизических систем. Для расширения сфер применения облакоцентричных архитектур в ориентации на киберфизические системы со сложными объектами и потенциально опасными технологическими процессами требуется их модернизация, направленная на частичную децентрализацию средств обработки информации.

2.3. Частичная децентрализация облачных систем

Полная централизация с режимами «тонких клиентов» не соответствует требованиям управления процессами функционирования распределённых киберфизических систем, в составе которых большое количество разнообразных объектов (датчиков, исполнительных устройств, контроллеров и другого оборудования), взаимодействующих через компьютерные сети. Необходимость непрерывной алгоритмической обработки потоков информации, сопровождающих их функционирование, диктуется требованиями и реального времени, и угрозами нештатных ситуаций, которые сопутствуют интенсивным, потенциально опасным режимам работы объектов высокой сложности.

Актуализация проблематики концепта Индустрия 4.0 потребовала дальнейшего развития облакоцентричных систем. С середины 10-х индустриально значимые свойства системной и вычислительной масштабируемости в изначально разнородных сетевых ресурсах достигаются в рамках растущего тренда, направленного на частичную децентрализацию серверных вычислительных ресурсов, направленную на переработку информации в реальном времени [8, 14]. Децентрализация – один из признаков стадии Web 3.0 (рис. 2).

В основу нового тренда развития положены требования к частичной децентрализации путём добавления иерархии дополнительных околооблачных сетевых слоёв периферийного окружения центрального облака. Дополнительные слои связывают объекты распределённых промышленных систем в функционально целостные производственные комплексы посредством аппаратно-программных интерфейсов сопряжения периферийных узлов с производственными объектами.

На рис. 4 приведена структура пост-облачной концепции частичной децентрализации. Она иллюстрируется на примере системы CDEF [18] с 4-х уровневой иерархией сетевых слоёв обработки данных: CC (Cloud Computing) – облачные вычисления, FC (Fog Computing) – туманные вычисления, EC (Edge Computing) – граничные вычисления, DC (Dew Computing) – росистые вычисления.

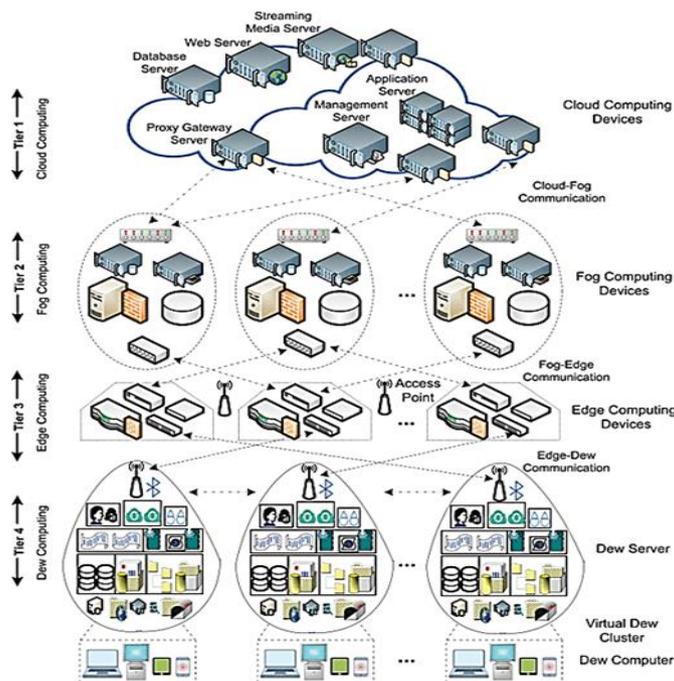


Рис. 4. Структура пост-облачной четырёхуровневой иерархии (<https://ieeexplore.ieee.org/document/8114187>)

В периферийных слоях осуществляется первичная обработка данных, формирование и передача агрегированной информации в облако, где происходит накопление, хранение и углублённая обработка. При этом сбор и первичный анализ данных проводится не в централизованной вычислительной среде, а в режимах «computing at the place» в тех местах (<https://community.hpe.com/t5/advancing-life-work/get-in-go-zoom-innovating-for-next-generation-automobile/ba-p/7087198>), где идёт генерация больших потоков данных (панели управления, датчики, отдалённые рабочие станции), там, где задержки носят критичный характер, а передача этих потоков в серверный центр на обработку в реальном времени не возможна. Тем самым снижается нагрузка на серверы (значит и стоимость центрального облака). Серверы во взаимодействии с периферийными сетевыми слоями осуществляют общий контроль и координацию всех технологических процессов. Особая роль при этом отводится технологиям ЦД. Они формируют «эталонную» среду штатного функционирования производственных систем с возможностями вывода оборудования из штатных/аварийных режимов. Одна из главных задач пост-облачной иерархии поддержание на регулярной основе и с необходимой эффективностью программирование и функционирование индустриально значимых ЦД на всех этапах жизненного цикла (см. рис. 1 и [4, 8, 14]).

Примеры применения частично децентрализованных систем – Интернет вещей (IoT), умные города, умные сети, виртуальная/дополненная реальность, беспилотные транспортные средства и другие сферы, охватываемые концептом Индустрия 4.0, который определяет стратегию цифровой трансформации систем производственной направленности в безопасно и эффективно функционирующие/развивающиеся киберфизические системы.

Частично децентрализованные системы выводят на рынки новое качество цифровых услуг в виде системно увязанного триплета <информация+алгоритмика+ЦД>, которое необходимо для осуществления этой стратегии. Технологии ЦД в отечественном исполнении обретают качественно

новые возможности для своего развития в качестве цифровой опоры для минимизации издержек в трансформации критически важных производств в киберфизические системы новых поколений.

Многослойная сателлитная децентрализация остаётся облакоцентричной, поскольку реализуема только с принципиальными ограничениями на объёмы периферийных расширений вычислительных ресурсов серверного центра. В таких решениях свойства системной целостности, а также системной и вычислительной масштабируемости достигаются путём многоуровневой стандартизации процессов формирования и развития экосистем производственной функционально-нишевой направленности.

3. Стадия гибридации серверных архитектур

В цифровой трансформации техногенных систем концепта Индустрия 4.0 частично децентрализованные пост-облачные системы выдвигаются на первые планы в направлениях рентабельного расширения сфер применения ЦД. При этом методы и технологии ЦД получают мощный импульс к дальнейшему своему развитию в части расширения структурного и вычислительного диапазонов масштабируемости моделей ЦД, реализуемых в облакоцентричных системах.

Современные облачные серверы – это высокопроизводительные универсальные компьютеры с многоядерными архитектурами. В серверах диапазоны масштабирования многоядерных вычислительных ресурсов, выделяемых для обработки потоков клиентских задач, не превышает нескольких десятков. Это является принципиальным ограничением вычислительной сложности для запросов, которым требуется больше ядер, чем в каждом из серверов. С ростом размеров и разнообразия больших систем при создании ЦД неизбежно возникает необходимость «тяжёлого» [4] моделирования сложных объектов, для которых требуются возможности кардинального (на порядки) увеличения диапазонов масштабирования вычислительной производительности на всех уровнях.

Такое увеличение достижимо путём гибридации серверов и сетевых узлов за счёт добавления к ним компьютеров-ускорителей с массовым многоядерным параллелизмом, в которых количество быстродействующих вычислительных ядер составляет многие сотни и тысячи. При этом диапазоны масштабирования вычислительной производительности вырастают на порядки, что открывает качественно новые возможности для развития технологий создания ЦД, направленных не только на углублённое моделирование всё более сложных компонентов больших распределённых систем, но и на повышение качества управления (в реальном времени) в таких системах в целом.

Развитие технологий ЦД, требующих существенного увеличения производительности, становится вполне возможным благодаря наличию весьма развитых технологий виртуализации облачных ресурсов. За два поколения своего развития они прошли апробацию огромной практикой успешного коммерческого продвижения разнообразных облачных систем на мировые рынки. К настоящему времени эти технологии позволяют (путём добавления интерфейсов и драйверов сопряжения) поддерживать широкое использование облачных систем в различных промышленных применениях.

Пост-облачные системы в критических сферах с повышенными требованиями к безопасности предполагают расширение сфер применения технологий ЦД с углублённым моделированием сложных узлов, агрегатов и систем, обеспечивающих достижение необходимых уровней безопасности. Далее покажем, как в новейших разработках, расширяющих возможности облакоцентричных систем, добавляются решения, направленные на виртуализацию вычислительных ресурсов с гибридными архитектурами.

3.1. GPU-as-a-Service: блоковая интеграция массового многоядерного параллелизма в виртуальную среду облакоцентричных систем

Фирма NVIDIA, формируя рыночную линейку общедоступных гибридных архитектур (как развитие видеокарт для ПК), в начале 10-х первой вывела на рынки однокристальные компьютеры-ускорители с массовым многоядерным параллелизмом общего назначения класса GP GPU (General-Purpose Graphics Processing Unit). В первом чипе такого ускорителя (с названием «Fermi»), изготовленного по технологии 40нм, было 512 универсально программируемых ядер. Благодаря прогрессу СБИС-технологий на технологиях 7 нм и менее количество таких ядер на кристалле увеличивается более, чем в 30 раз.

Чипы ускорителей широко используются в составе гибридных компьютерных систем различных классов от ПК и встраиваемых систем до мощных серверов и суперкомпьютеров. При этом универсальный компьютер (хост-машина) исполняет системную роль «ведущего» агрегата, а присоединяемый ускоритель – «ведомого». Производительность таких систем на задачах с массовым параллелизмом может на порядки превышать предельную производительность хост-машин.

Новейшие исследования фирмы NVIDIA обозначили одно из главных направлений продвижения облакоцентричных систем в индустриальные сферы, в которых в текущих и перспективных разработках требуются повышение вычислительной производительности на порядки. Исследования направлены на гибридизацию серверных архитектур облачных систем путём виртуального сопряжения уже виртуализированных серверных ресурсов применительно к ускоряющим блокам видеокарт A100 (рис. 5а), построенными по технологии GP GPU.

В ходе этих исследований осуществляются системотехнические разработки (<https://object-storage-ca-ymq-1.vexxhost.net/swift/v1/6e4619c416ff4bd19e1c087f27a43eea/www-assets-prod/presentation-media/OpenStack-Boston-Summit-Presentation.pdf>), направленные на интеграцию продукции NVIDIA в существующие среды виртуализации ресурсов облачных систем. Это опережающий шаг продвижения однокристалльных компьютеров-ускорителей с массовым многоядерным параллелизмом в новые сферы разнообразных применений ЦД путём расширения пакета комплекса взаимосвязанных облачных услуг до системного триплета «информация+алгоритмика+модели ЦД».

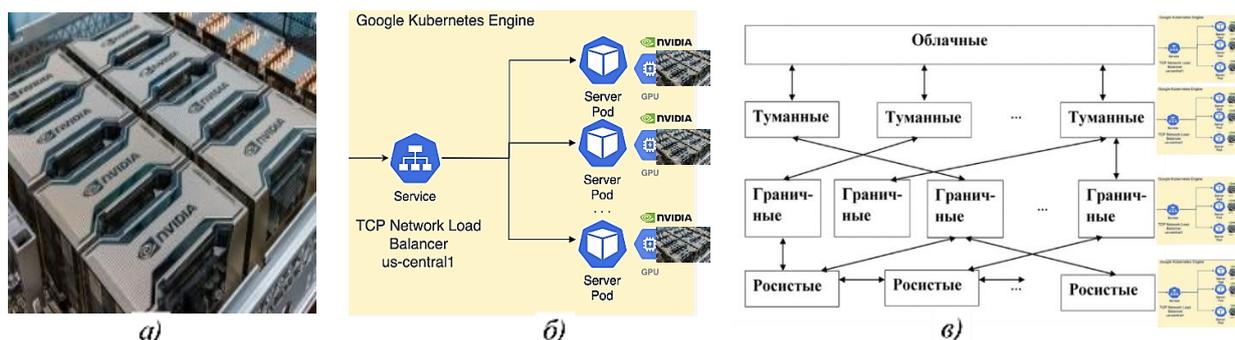


Рис. 5. GPU-as-a-Service: а) блок ускорителей NVIDIA A100 на базе чипа GP GPU Ampere, б) вычислительная масштабируемость на уровне многоядерных блоков A100 посредством контейнеризации в виртуальной среде облачных систем, в) гибридизация компьютерных устройств на разных уровнях путём их сопряжения с блоками A100

Новизна этого подхода – сопряжение блоков ускорителей с сервером не посредством жёстко фиксированной системотехнической схемы, а посредством интеграции виртуализированных серверных ресурсов с виртуализированными ресурсами ускорительных блоков (рис. 5б). Виртуальная среда позволяет динамически формировать программно управляемые конфигурации используемых ресурсов в зависимости от запросов со стороны исполняемых задач, что является основным режимом облачных систем по обслуживанию потока запросов на информационно-алгоритмические услуги.

На рис. 5в показаны возможности интеграции блоковых ускорителей GP GPU с ресурсами разных уровней пост-облачных систем, для чего требуется формирование единой среды виртуализации ресурсов не только в серверном центре, но и с охватом всех около серверных уровней.

Сопряжение в единой виртуальной среде позволяет проводить масштабирование массового многоядерного параллелизма в виртуальной среде пост-облачных систем в режимах «по требованию». Это означает, что применение в составе облачных ресурсов компьютеров-ускорителей с тысячами вычислительных ядер позволяет при обработке потока запросов расширять диапазон масштабирования вычислительной производительности на порядки по сравнению с имеющимися возможностями многоядерных универсальных компьютеров, в которых предельное число ядер имеет непреодолимое их ограничение несколькими десятками.

В этом принципиальное отличие виртуализированной гибридизации компьютерных средств в составе облакоцентричной системы.

3.2. PS(П)C)-as-a-Service: внутричиповая интеграция массового многоядерного параллелизма в виртуальную среду облакоцентричных систем

В ориентации на развитие суверенных облакоцентричных технологий и технологий ЦД рассмотрим возможности импортонезависимого привнесения [4] системных функций внутричиповой виртуализации массового многоядерного параллелизма применительно к перспективным отечественным однокристалльным компьютерам-ускорителям ПС-2000М, в основу которой положена масштабируемая многопроцессорная SIMD-архитектура отечественного компьютера-ускорителя ПС-2000 (рис. 6а). Высокопроизводительные гибридные ВК на базе ПС-2000 [19] выпускались большой

промышленной серией в отсутствие мировых аналогов. Высокая эффективность этой архитектуры подтверждена широкой практикой применения в различных сферах народного хозяйства.

На рис. 6б в обобщённом виде представлена структура чипа PS-2000M [19], составленная из базовых вычислительных модулей (БВМ). В их основе апробированная практикой архитектура PS-2000. Одно из ключевых преимуществ архитектуры PS-2000M состоит в её структурной масштабируемости. Данное качество позволяет эффективно наращивать производительность однокристалльного ускорителя по мере увеличения количества вычислительных ядер, размещаемых на кристалле ускорителя. Такое масштабирование производительности становится возможным, начиная с доступной в стране технологии 90 нм (256/512 ядер на кристалле [20]). На передовых технологиях (менее 10 нм) число ядер в чипе может достигать тысяч и десятков тысяч ядер [19].

В отличие от подхода NVIDIA к облачной виртуализации многоядерного параллелизма ускорителей GP GPU (GPU-as-a-Service) облачная виртуализация структурно-масштабируемого параллелизма ускорителей PS(ПC)-2000M предполагает переход на более глубокие уровни архитектурного погружения в ресурсы массового многоядерного параллелизма. Если виртуализация GPU-as-a-Service в архитектуре GP GPU осуществляется на блоковом уровне из видеокарт (рис. 5), то в нашем случае виртуализация PS(ПC)-as-a-Service воплощается на внутричиповом уровне над множеством БВМ (рис. 6б), а в перспективе с включением в пространство виртуализации вычислительных ядер, из которых составлены все БВМ.

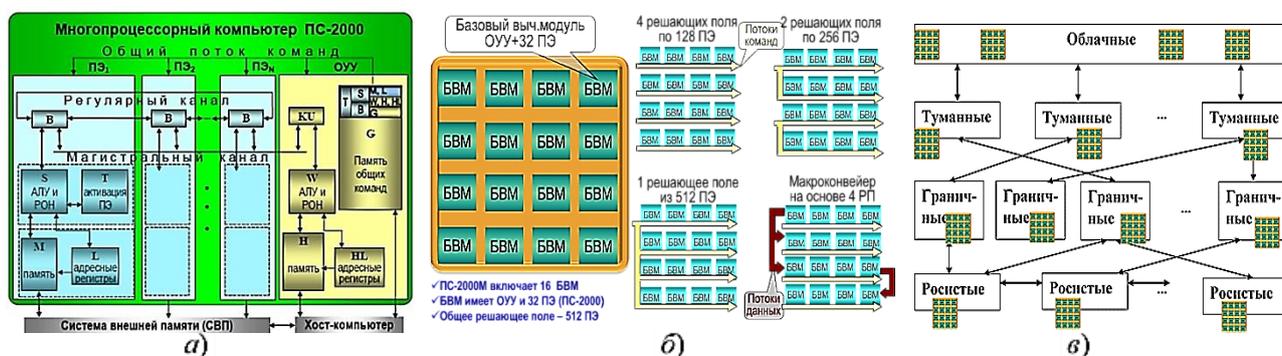


Рис. 6. Многопроцессорный компьютер-ускоритель: а) структура PS-2000, б) PS-2000M (от 90-45нм и менее, проект). Однокристалльный компьютер-ускоритель с внутрикристалльной программной реконфигурацией, в) PS(ПC)-as-a-Service (проект)

На рис. 6в показана возможность гибридации серверов и компьютерных устройств в периферийных сетевых узлах, для которых требуется кардинальное расширение диапазона производительности путём масштабирования массового внутрикристалльного многоядерного параллелизма компьютеров-ускорителей PS-2000M.

В составе виртуальной среды импортонезависимых облакоцентричных систем (с серверами на базе отечественных микропроцессоров) такое решение позволит на мировом концептуальном уровне облачного прогресса расширять диапазоны масштабирования вычислительной производительности.

4. Заключение

Для критически важных распределённых техногенных систем обозначены направления и пути импортонезависимого развития больших компьютерно-сетевых вычислительных сред с применением отечественной элементной базы и методов реализации в них цифровых двойников (ЦД).

В рассмотрении трёх стадий эволюционного развития ГКС проведён анализ общесистемных и системотехнических аспектов развития системообразующих и функциональных возможностей клиент-серверных сетевых инфраструктур. На этой основе объясняются особенности формирования глобальных цифровых услуг и рынков массового производства-потребления гипертекстового информационного контента (Web-технологии), затем информационно-алгоритмических услуг «в одном пакете» (облачные технологии) и, наконец, комплексных цифровых услуг в виде триплета <информация+алгоритмика+модели ЦД> с системно интегрированными в рамках экосистем решениями на основе многослойных пост-облачных технологий. В условиях глобальной информационной сильносвязности обеспечение безопасно-устойчивого функционирования/развития сложных промышленных систем концепта Индустрия 4.0 без массового применения технологий ЦД не представляется возможным.

Цель дальнейших исследований, развиваемых в работах [1-3, 16, 17, 19, 20], – выявление и применение базовых закономерностей развития ГКС для формирования в сколь угодно больших сетях универсального бесшовно программируемого и кибербезопасного алгоритмического пространства распределённых и параллельных вычислений [2, 3], в котором снимаются ограничения на масштабы реализуемых ЦД, свойственные разрозненным экосистемам облакоцентричных систем.

Литература

1. *Затуливетер Ю.С., Фищенко Е.А.* Проблемы программируемости, безопасности и надежности распределенных вычислений и сетецентрического управления. Ч. 1. Анализ проблематики // Проблемы управления. 2016. N 3. – С. 49-57.
2. *Затуливетер Ю.С.* Проблемы глобализации парадигмы управления в математически однородном поле компьютерной информации // Проблемы управления. 2005. N 1. Ч. I. – С. 1-12; N 2. Ч. II. – С. 13-23.
3. *Затуливетер Ю.С., Фищенко Е.А.* К компьютерно-сетевым архитектурам для цифровой трансформации больших систем // Программные системы: теория и приложения. 2020. т. 11, N 3(46). – С. 85–131.
4. *Байдоров Д.Ю., Абакумов Е.М., Файков Д.Ю.* Программное обеспечение «тяжелого» класса: возможности импортозамещения // Вопросы инновационной экономики. 2022. Том 12. N 1. – С. 295-316. doi: 10.18334/vinec.12.1.114143. URL: https://www.elibrary.ru/download/elibrary_48333791_94658944.pdf, <https://1economic.ru/lib/114143>.
5. *Полетыкин А.Г., Промыслов В.Г., Семенов К.В., Менгазетдинов Н.Э., Жарко Е.Ф.* Применение цифрового двойника в жизненном цикле АСУ ТП АЭС // Автоматизированные системы управления технологическими процессами АЭС и ТЭС: материалы II Международной научно-технической конференции. – Минск: БГУИР, 2021. – С. 193-197.
6. *Панпе И.Н., Володин В.С.* Применение математических моделей на различных стадиях жизненного цикла российских АСУ ТП АЭС// Автоматизированные системы управления технологическими процессами АЭС и ТЭС: материалы II Международной научно-технической конференции. – Минск: БГУИР, 2021. – С. 182-188.
7. *Дорохова И.* Заглянуть в будущее. Виртуально-цифровая АЭС помогает прогнозировать поведение ее реального прототипа // Вестник атомпрома. 2021. N 10. – С.14-18. URL: https://atomvestnik.ru/wp-content/uploads/2021/12/VA10_Book_spread.pdf, <https://atomvestnik.ru/2021/12/22/zagljanut-v-budushhee/>.
8. *Прохоров А.Н., Лысачев М.Н.* Цифровой двойник. Анализ, тренды, мировой опыт. – М.: ООО «АльянсПринт», 2020. – 401 с.
9. *Полетыкин А.Г.* Элемент цифровой индустрии 4.0: виртуальная суперкомпьютерная модель для сопровождения и модернизации сложных систем // Управление развитием крупномасштабных систем (MLSD'2018): материалы Одиннадцатой междунар. Конфер. – М.: ИПУ РАН, 2018. Т. 1. – С. 54-59.
10. *Жарко Е.Ф.* Цифровой двойник технологического процесса энергоблока АЭС // Управление развитием крупномасштабных систем (MLSD'2022): труды Пятнадцатой международной конференции. – М.: ИПУ РАН, 2022. – С. 1053-1059.
11. "Росатом" разработает цифровые двойники АЭС малой мощности - Техэксперт / URL: <https://cntd.ru/news/read/rosatom-razrabotaet-cifrovye-dvoyniki-as-maloy-moschnosti>.
12. *Плотонова О.* Росатом создает виртуальную АЭС // Атомный эксперт, 2018. N 1 (62). – С.20–27. URL: https://archive.atomicexpert.com/virtual_npp_rosatom_.
13. *Промыслов В.Г., Полетыкин А.Г., Семенов К.В., Менгазетдинов Н.Э., Бывайков М.Е., Степанов В.Н.* Технология распределенной разработки ПО для АСУ ТП АЭС с использованием виртуализации и цифровых двойников // Управление развитием крупномасштабных систем (MLSD'2021): труды Четырнадцатой международной конференции. – М.: ИПУ РАН, 2021. – С. 98-102.
14. *Прохоров А., Коник Л.* Цифровая трансформация. Анализ, тренды, мировой опыт. Издание второе, исправленное и дополненное. — М.: ООО «КомНьюс Групп», 2019. — 368 с.
15. *Прохоров А., Рахматуллин С.* Центры обработки данных. Анализ, тренды, мировой опыт. – М.: ООО «АльянсПринт», 2021. – 416 с.
16. *Затуливетер Ю.С.* Предпосылки появления и пути реализации глобального управляющего компьютера // Датчики и системы. 2005. N 4. –С. 47-58; N 5. – С. 25-33.
17. *Затуливетер Ю.С., Фищенко Е.А.* Синергия цифровой универсальности глобальной компьютерной среды // Программная инженерия. 2022. N 3. – С. 107 - 118.
18. *Ray P. P.* An Introduction to Dew Computing: Definition, Concept and Implications // IEEE Access, Vol. 62017. DOI: 10.1109/ACCEDESS.2017.2775042, pp. 723-737.
19. *Затуливетер Ю.С., Фищенко Е.А., Артамонов С.Е., Козлов В.А.* Элементы стратегии и архитектурные предпосылки опережения в области однокристалльных многопроцессорных компьютеров с массовым параллелизмом // Информационные технологии. 2014. N 2. Приложение. – С. 1-32.
20. *Затуливетер Ю.С., Фищенко Е.А.* Развитие импортонезависимой элементной базы для поддержки технологий «цифровых двойников» (на примере задач АСУ ТП АЭС) // Управление развитием крупномасштабных систем (MLSD'2022): труды Пятнадцатой международной конференции. – М.: ИПУ РАН, 2022. – С. 1060-1067.