DOI: 10.18372/2410-7840.20.13426
УДК 004.274:004.056

ЦЕНТРАЛИЗОВАННЫЙ СИНТЕЗ РЕКОНФИГУРИРУЕМЫХ АППАРАТНЫХ СРЕДСТВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ НА ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ПЛАТФОРМАХ

Виктор Евдокимов, Анатолий Давыденко, Сергей Гильгурт

Основная функция сигнатурных сетевых систем обнаружения вторжений (ССОВ) – поиск в интенсивном потоке данных признаков известных атак из базы сигнатур, содержащих на сегодняшний день десятки тысяч записей. В отличие от межсетевого экрана ССОВ обрабатывает не только заголовки, но и тела пакетов, осуществляя так называемый глубокий анализ пакетов – DPI. Задача множественного распознавания строк – особый тип распознавания, при котором исходный текст анализируется одновременно на наличие множества паттернов. Из-за стагнации частоты микропроцессоров, а также в связи с неизменным ростом сетевого трафика, увеличением количества и сложности атак традиционным программным решениям все сложнее отвечать ужесточающимся требованиям информационной безопасности. В этой связи все большее распространение получают аппаратные решения с применением реконфигурируемых устройств на базе ПЛИС типа FPGA, которые сочетают в себе близкую к аппаратной производительность с гибкостью программного обеспечения. К сожалению, разработка сложных реконфигурируемых устройств является нетривиальной задачей. Пользователи ССОВ, которыми обычно являются системные администраторы, не имеют для этого ни достаточной квалификации, ни требуемых вычислительных ресурсов. С другой стороны, специфика задач информационной безопасности требует частого выполнения процедуры пересинтезирования реконфигурируемых ускорителей. Для решения проблемы было предложено централизовать вычислительный процесс с использованием грид-инфраструктуры и облачной платформы. Такой подход позволяет перенести трудоемкую и вычислительно сложную процедуру из локальных вычислительных сетей в высокопроизводительную среду. Был разработан и протестирован экспериментальный сервис. Приведены первые полученные результаты. Проведено предварительное сравнение грид- и облачной технологий. Помимо кибербезопасности, ускорение задачи множественного распознавания актуально и для многих других важных приложений, таких как интеллектуальный анализ данных (data mining), ускоренная обработка XML-запросов, управление технологией QoS, фильтрация в IP-телефонии, оптимизация кеширования и др.

Ключевые слова: *информационная безопасность, ССОВ, глубокий анализ пакетов, множественное распознавание, ПЛИС, централизованный синтез, высокопроизводительные вычисления, грид, облачные вычисления.*

Введение

Все возрастающая в последние годы интенсивность и изощренность вредоносной активности в информационно-вычислительных системах вынуждают совершенствовать и развивать как организационные, так и технические решения в области киберзащиты. К последним относятся системы обнаружения вторжений (СОВ), предназначенные для мониторинга злонамеренной активности и выявления атакующих действий со стороны нарушителей [1]. Поскольку такие системы ориентированы преимущественно на внешне угрозы, реализуемых по каналам передачи данных, подразумеваются чаще всего сетевые системы обнаружения вторжений (ССОВ), соответствующий англоязычный термин – Network Intrusion Detection System (NIDS). Переменчивая природа киберугроз вынуждает искать технические решения, способные противостоять как известным, так и новым либо модифицированным атакам, свойства которых неизвестны или недостаточно изучены. Как правило, подобные системы основаны на анализе аномального поведения субъектов сетевой инфраструктуры. К сожалению, решениям подобного плана все еще свойственны серьезные недостатки, такие как высокая интенсивность ложных срабатываний, сложность процесса настройки, временная затратность на обучение и создание профилей нормального состояния [2]. Поэтому в реальных разработках на сегодняшний день чаще используют системы, основанные на анализе описаний известных атак в виде сигнатур (Signature-based NIDS).

Необходимость распознавания в интенсивном потоке сетевых пакетов признаков большого числа известных атак при неизменно увеличивающихся объемах, передаваемых данных и остановившемся росте частоты микропроцессорных систем, вынуждает разработчиков отказываться от программных решений и ориентироваться на аппаратные. Основное внимание при этом уделяется реконфигурируемым устройствам на базе программируемых логических интегральных схем (ПЛИС) [3]. Близкое к аппаратному быстродей-

ствие программируемой логики в сочетании с высокой гибкостью наиболее полно отвечает требованиям такой динамично развивающейся области, как защита информации. Следует заметить, что сигнатурный метод распознавания признаков злонамеренной активности также успешно используется при создании на базе ПЛИС антивирусов, систем противодействия сетевым червям и других реконфигурируемых аппаратных средств информационной безопасности (РАСИБ).

Широкое применение реконфигурируемых ускорителей, однако, сдерживается рядом факторов. Процесс разработки аппаратного устройства на базе ПЛИС является сложной задачей, требующей, с одной стороны, интеллектуального труда квалифицированных специалистов, с другой – вычислительных затрат на генерирование так называемой конфигурации – последовательности битов, которую необходимо загрузить в ПЛИС перед ее использованием для задания внутренних соединений с целью придания изделию требуемой функциональности. Пользователи средств информационной безопасности не являются специалистами в области реконфигурируемых средств и не обладают ни навыками разработки аппаратных ССОВ, ни достаточными вычислительными ресурсами. **Целью настоящей работы** является всесторонний анализ предлагаемого решения указанной проблемы, заключающегося в организации централизованной системы синтеза РАСИБ на базе высокопроизводительных вычислительных платформ, а именно – грид-инфраструктуры и облачного сервиса.

1. Аппаратная реализация систем обнаружения вторжений

1.1. Анализ сетевых систем обнаружения вторжений. Исторически первыми разработками в сфере защиты информации, ориентированными на применение ПЛИС типа Field Programmable Gate Array (FPGA), были сетевые системы обнаружения вторжений. В публикации [4] проведено исследование и обобщение основных принципов построения систем обнаружения вторжений на базе программируемой логики по результатам анализа накопленного в мире опыта построения ССОВ на базе ПЛИС типа FPGA.

Система обнаружения вторжений в обязательном порядке включает в себя один или несколько сенсоров и интерфейс с пользователем. [1].

В зависимости от защищаемого объекта, различают системы обнаружения вторжений:

– контролирующие отдельные компьютеры;

– анализирующие пакеты сетевого трафика всей локальной сети.

Наибольший эффект от применения аппаратного решения достигается в сетевых системах обнаружения вторжений – ССОВ. В настоящей работе речь пойдет именно о таких системах.

Механизм функционирования сигнатурной сетевой системы обнаружения вторжений в общем случае состоит из 3-х этапов:

– захват сетевых пакетов (packet capture);

– фильтрация и сборка пакетов (filtering / fragmentation reassembly);

– распознавание (pattern matching).

Самым ресурсоемким является последний этап, который сводится к выполнению большого объема работы по анализу содержимого сетевых пакетов на предмет совпадения с паттернами – последовательностями символов из базы данных сигнатур.

Анализ сетевого трафика может осуществляться двумя способами:

– путем тотального захвата и инспектирования всех (необработанных – raw) пакетов сетевого трафика;

– с учетом сетевых протоколов – stateful подход, основанный на анализе заголовков сетевых пакетов с целью полного восстановления сеансов сетевого обмена.

Системы, основанные на первом способе, распознают большее число атак. Для них не являются проблемой нестандартные номера портов, потерянные либо намеренно искаженные злоумышленником сетевые пакеты. С другой стороны, такие ССОВ намного более ресурсоемки в своей реализации.

1.2. Некоммерческие базы данных сигнатур для СОВ. В большинстве научных исследований по реконфигурируемым ССОВ в качестве базы данных сигнатур, которые должны распознаваться создаваемой системой защиты, используются наборы из свободно распространяемых систем обнаружения вторжений, таких как *Bro, Hogwash, Snort* либо *Suricata.* Предпочтение при этом в большинстве случаев исследователи отдают системе Snort как устоявшемуся де-факто стандарту [5]. Подобный подход позволяет, с одной стороны, приблизить проводимые исследования к реальной жизни, с другой – использование одинаковых наборов тестовых данных позволяет более корректно сравнивать различные методы, алгоритмы и решения для ССОВ, объективно оценивать полученные технические показатели.

В работе [6] подробно рассмотрена структура записей базы данных сигнатур ССОВ Snort. Здесь под понятием *сигнатура* понимается вся совокупность информации о конкретной угрозе, содержащаяся в одной записи базы данных Snort. В этом смысле термин «сигнатура» эквивалентен понятию записи базы данных. Уточним также, что под словом *паттерн* мы понимаем образец текстовой строки (фиксированную последовательность символов, закодированную байтами), входящий в состав сигнатуры, который ищется в теле анализируемого сетевого пакета. В общем случае сигнатура содержит один или несколько паттернов, а также вспомогательные правила, регламентирующие условия их поиска. По данным правилам обрабатываются заголовки сетевых пакетов при анализе трафика в режиме учета сетевых протоколов (см. п. 1.1).

2. Задача множественного распознавания строк

2.1. Распространенные алгоритмы распознавания строк. Основная техническая задача, решаемая при сигнатурном анализе, в терминах теории вычислений на строках формулируется следующим образом. Необходимо найти все вхождения некоторой заданной строки символов (подстроки) в более длинной строке (тексте). Искомую строку называют паттерном [7].

Данная задача давно и успешно решается в компьютерных приложениях. Известно множество алгоритмов ее решения, таких как алгоритмы Кнута–Морриса–Пратта, Бойера–Мура, Карпа-Рабина и их многочисленные модификации [7]. Однако большинство из них ориентировано на программную (последовательную) реализацию в однопроцессорных системах. Но, что более существенно, данные алгоритмы являются по своей сути одношаблонными, то есть ориентированными на распознавание одной подстроки в отдельный момент времени. Последовательное применение таких алгоритмов по очереди для каждого паттерна оказывается неэффективным по причинам, указанным ниже.

2.2. Специфика распознавания строк в сигнатурных системах распознавания. Задаче распознавания строк применительно к системам распознавания сигнатур на базе РАСИБ в значительной степени свойственен параллелизм, причем, по двум направлениям: во-первых, несколько сетевых пакетов могут анализироваться одновременно; во-вторых, сравнение может производиться сразу со многими подстроками из базы данных сигнатур. Рассмотрим эти направления.

При реализации параллелизма первого типа возникает трудноразрешимое противоречие: разделение интенсивного входного потока на большое число отдельных блоков, обрабатываемых независимыми вычислительными модулями, приводит к задержкам, пропорциональным коэффициенту распараллеливания, обусловленным большим размером блоков; уменьшение же их размера снижает полезный эффект от распараллеливания из-за вынужденного перекрытия, тем большего, чем длиннее искомые подстроки. К тому же, такой подход отличается сложностью реализации из-за необходимости решения вспомогательных задач управления, диспетчеризации и буферизации [8].

Распараллеливание второго типа – по подстрокам, то есть, разделение на подгруппы набора распознаваемых образцов, также возможно. Но в этом случае обнаруживается важная особенность – сигнатуры в базе данных во многом повторяют друг друга. Причем, данное свойство самоподобия в силу конечности алфавита, теоретически, должно усиливаться по мере увеличения размера базы данных сигнатур. Учет данного эффекта позволяет существенно повысить производительность распознающей системы. Упомянутые выше одношаблонные алгоритмы в этом отношении оказываются неэффективными.

2.3. Формулировка задачи множественного распознавания. Резюмируя сказанное выше, приходим к заключению, что в сигнатурных системах распознавания речь идет о несколько иной задаче, нежели одношаблонный поиск подстроки в строке. В дальнейшем будем называть ее задачей множественного распознавания строк (Multi-Pattern String Matching) [8]. Суть данной задачи заключается в необходимости быстрого одновременного поиска во входной последовательности символов всех паттернов, входящих в заданный набор подстрок (словарь), при этом различные фрагменты подстрок, входящих в словарь, в значительной степени повторяют друг друга. Важно также отметить, что при решении данной задачи должен максимально использоваться естественный параллелизм, как по входным данным, так и по распознаваемым подстрокам.

2.4. Алгоритмы множественного распознавания. Наиболее распространенным решением данной задачи считается алгоритм Ахо-Корасик [9], послуживший прототипом для огромного числа модификаций и доработок, большинство из которых ориентировано на аппаратную реализацию. Его суть заключается в построении по определенным принципам цифрового автомата, на вход которого в процессе распознавания последовательно, символ за символом, поступает

информация. В зависимости от содержимого входных данных автомат переходит в состояния, сигнализирующие о наличии в них той или иной эталонной строки. Таким образом, набор распознаваемых образцов оказывается аппаратно "вшитым" в структуру распознающего устройства, реализующего цифровой автомат.

Как будет показано ниже в разделе 4.4, при создании аппаратных распознающих схем цифровые автоматы являются далеко не единственно возможным подходом.

3. Параметры и требования, предъявляемые к ССОВ на ПЛИС

Сформулируем требования, предъявляемые к системам обнаружения вторжений и другим средствам киберзащиты на базе реконфигурируемых аппаратных ускорителей, а также основные параметры, по которым следует оценивать их эффективность, исходя из рассмотренных выше особенностей задачи множественного распознавания, которая должна в них решаться [8].

Главными показателями производительности ССОВ являются *максимальное число сигнатур*, распознаваемых системой, и *пропускная способность*, которая может при этом быть достигнута.

Однако на практике более важной и труднодостижимой характеристикой ССОВ является *масштабируемость* – способность наращивать возможности в широких пределах без несоизмеримо высоких дополнительных затрат. Актуальность данного свойства технического решения для сетевой системы обнаружения вторжений обусловлена, с одной стороны, стремительным ростом сетевого трафика, с другой – постоянным увеличением размеров базы данных сигнатур. В этой связи следует разделять два направления приложения данного показателя – по быстродействию и по числу сигнатур.

Важным показателем ССОВ, также непосредственно связанным с производительностью, является *предсказуемость пропускной способности*, то есть, независимость ее временных характеристик от состава входных данных. Обнаружение злонамеренного контента в сетевом трафике является редким событием, вероятность возникновения которого в штатном режиме невелика. Но если содержимое анализируемых сетевых пакетов существенно влияет на быстродействие модуля распознавания ССОВ, такая система может оказаться уязвимой к намеренному засорению сетевого трафика сигнатурами известных атак злоумышленником.

Специфической чертой систем обнаружения вторжений на базе сигнатур является необходимость регулярного обновления активной базы данных. Возможности *динамического обновления* существенно влияют на практическую полезность технического решения. Данный показатель затрагивает такие моменты, как возможность обновления базы сигнатур без приостановки процесса распознавания, способность обходиться без перепрограммирования ПЛИС, либо, в противном случае, наличие средств автоматической генерации и загрузки в ПЛИС новой конфигурации, а также удобство и скорость выполнения данной операции.

Независимость от состава сигнатур также является важной характеристикой ССОВ. Ориентация модуля распознавания на ограниченный алфавит с целью повышения быстродействия может привести к нежелательным последствиям при его использовании в системах защиты информации.

Помимо скоростных характеристик систем обнаружения вторжений, для их практического использования важны также стоимостные показатели. *Объем оперативной памяти*, необходимой для реализации выбранного алгоритма распознавания, существенно влияет, в итоге, на быстродействие. Если имеющихся в ПЛИС ресурсов быстродействующей блочной памяти (BRAM) не достаточно для реализации запоминающего устройства, то возникает необходимость во внешней памяти, которая намного медленнее внутренней.

Аппаратные затраты при работе с ПЛИС принято оценивать площадью кристалла, задействованной для реализации конкретного устройства, абсолютной или в процентной доле от общего ресурса микросхемы. В качестве единиц измерения данного показателя используют либо условные эквивалентные логические элементы (вентили – калька из англоязычной терминологии), либо системные логические элементы, либо другие структурные компоненты ПЛИС конкретного семейства конкретного производителя – поисковые таблицы (LUT), конфигурируемые логические блоки, секции и т.п. [10].

Существенной является также *общая стоимость реализации системы*. Каким бы эффективным ни был распознающий модуль, если для его интеграции в ССОВ необходимы существенные дополнительные затраты, например, на преобразование формы представления информации, общая стоимость решения может оказаться неприемлемой.

4. Основы построения ССОВ на реконфигурируемой платформе

4.1. ССОВ и файерволы. Сетевые системы обнаружения вторжений, используются для защиты локальных вычислительных сетей, при этом они решают задачи, дополняющие функции файервола (межсетевого экрана). Подавляющее число несанкционированных проникновений в защищаемую сеть может быть отфильтровано по адресной информации, содержащейся в заголовках сетевых пакетов, чем и занимается файервол. Однако в ряде случаев злоумышленнику удается обойти эту защиту и проникнуть в охраняемую локальную сеть. Именно для выявления таких ситуаций предназначены ССОВ.

4.2. Глубокий анализ пакетов. Системы обнаружения вторжений сигнатурного типа анализируют не только заголовки, но и тела сетевых пакетов, то есть, выполняют так называемую глубокую обработку пакетов (DPI – Deep Packet Inspection). При этом решается ресурсоемкая задача множественного распознавания, сложность которой постоянно растет. По этой причине в последнее время разработчики таких систем все больше внимания уделяют реконфигурируемым устройствам на базе ПЛИС типа FPGA [11]. Аппаратное быстродействие программируемой логики и построенных на ее базе реконфигурируемых вычислителей [12] в сочетании с их высокой гибкостью позволяют эффективно использовать естественный параллелизм, присущий задаче множественного распознавания в приложениях информационной защиты [13].

4.3. Обобщенная структура СОВ на базе ПЛИС. Сформулируем в общем виде состав и структуру аппаратной реализации на ПЛИС системы обнаружения вторжений, отвечающую сформулированным выше требованиям. На рис. 1 приведена обобщенная структурная схема получившегося решения.

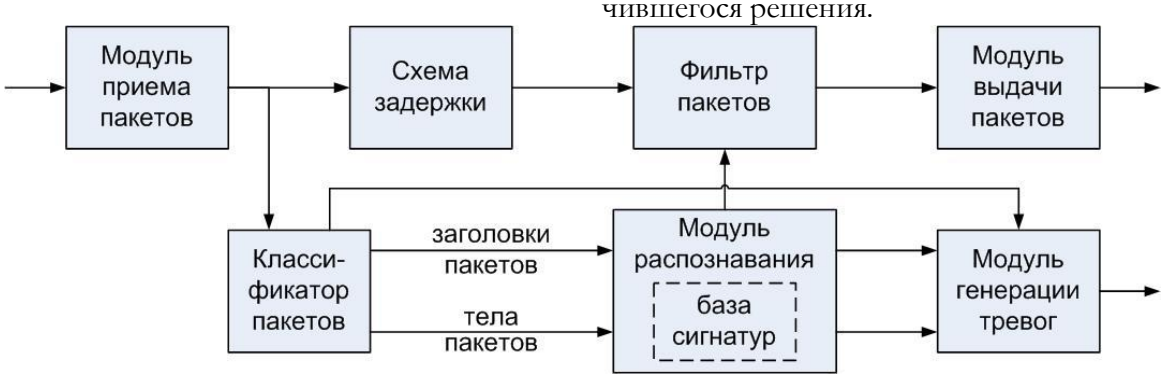

Рис. 1. Обобщенная структурная схема ССОВ на базе ПЛИС

В состав структуры входят:
– модуль приема пакетов;
– классификатор пакетов;
– модуль распознавания;
– модуль генерации тревог;
– схема задержки;
– фильтр пакетов;
– модуль выдачи пакетов.

Модуль приема пакетов осуществляет низкоуровневый захват сетевых пакетов и их преобразование в более удобный для внутрисхемной обработки тип кодирования, например, из формата XAUI (10 Gigabit Attachment Unit Interface) в формат XGMII (10 Gigabit Media Independent Interface).

Классификатор разбирает пакеты на основе анализа заголовков вплоть до требуемого уровня в зависимости от используемого метода обнаружения вторжений [8].

Модуль распознавания выполняет самую ресурсоемкую вычислительную операцию поиска сигнатур в соответствии с выбранным алгоритмом. От качества его реализации в значительной степени зависят главные характеристики всей системы обнаружения вторжений: производительность, ресурсоемкость и масштабируемость. В большинстве случаев база данных правил распознавания сигнатур, включая сами сигнатуры непосредственно "вшита" в распознающую вычислительную структуру. Вопросы выбора алгоритма распознавания и технического решения для его реализации выходят за рамки данной работы и будут рассмотрены в последующих публикациях цикла. Помимо аппаратуры, реализующей собственно алгоритм поиска вхождения шаблонов в содержимом сетевых пакетов, модуль распознавания в общем случае содержит также схему распознавания заголовков пакетов и детектор правил базы данных сигнатур.

Модуль генерации тревог служит для формирования сообщений об обнаруженных вторжениях. Он идентифицирует вредоносные пакеты и

объединяет информацию о типе атаки, поступающую от узла распознавания с дополнительными сведениями из заголовков пакетов, позволяющими идентифицировать источник вторжения.

Схема задержки синхронизирует поток пакетов с работой модуля распознавания.

Фильтр пакетов служит для пресечения вредоносного трафика путем отбрасывания вредоносных пакетов.

Модуль выдачи пакетов реализует преобразование, обратное тому, что выполнялось в модуле приема пакетов.

Следует заметить, что последние три компонента из рассмотренных, а именно: схема задержки, фильтр пакетов и модуль выдачи пакетов, в общем случае не являются необходимыми компонентами ССОВ и присутствуют только в сетевых системах предотвращения вторжений (ССПВ), соответствующий англоязычный термин – Network Intrusion Prevention System (NIPS). ССПВ предъявляют повышенные требования, как по быстродействию, так и по достоверности процедуры распознавания, поскольку могут оказывать существенное, в том числе и негативное влияние на работу защищаемой вычислительной сети. Системы предотвращения вторжений в данном исследовании не рассматриваются, а соответствующие им компоненты приведены лишь для полноты изложения.

4.4. Подходы к построению модуля распознавания ССОВ. Как указывалось выше, модуль распознавания является наиболее важным компонентом РАСИБ, от параметров которого во многом зависит эффективность средства информационной защиты в целом. Анализ информационных источников показывает, что в существующих на сегодняшний день системах аппаратного распознавания применяются самые разнообразные подходы, приемы и технические решения. Наиболее распространенными среди них являются:

– ассоциативная память на параллельных дискретных компараторах и ее разновидности;

– схемы на базе хэш-функций, в частности, фильтр Блума;

– конечные автоматы (finite automaton), реализующие в большинстве случаев алгоритм Ахо-Корасик [8].

Сравнительный анализ перечисленных подходов приводит к выводу, ни один из подходов не демонстрирует явных преимуществ перед другими и не удовлетворяет в полной мере требованиям, предъявляемым сетевыми системами обнаружения вторжений к их реализации на реконфигурируемых устройствах.

Так, параллельные компараторы и построенные на их основе разновидности ассоциативной памяти обеспечивают максимальное быстродействие, но дороже других подходов в плане потребления аппаратных ресурсов и электроэнергии, а также проигрывают в плане масштабирования. Фильтр Блума более экономичен и лучше масштабируется, но накладывает ограничение по длине сигнатур, а также требует дополнительных затрат на доуточнение результатов совпадения из-за свойственных ему по определению ошибок распознавания второго рода. Конечные автоматы обеспечивают стабильную, но относительно невысокую пропускную способность, сложны в построении и конфигурировании, требуют "взрывообразно" много памяти для больших словарей сигнатур.

5. Централизация синтеза реконфигурируемых аппаратных средств информационной безопасности

Сложности создания реконфигурируемых систем сдерживают быстрое распространение программируемой логики, в том числе – в сфере информационной безопасности. В данном разделе в качестве предлагаемого решения рассматривается подход, основанный на централизации процесса синтеза реконфигурируемых цифровых схем.

5.1. Проблемы синтеза аппаратных ускорителей для информационной безопасности. Рассмотрим сложности, которые возникают при создании и эксплуатации реконфигурируемых средств аппаратного ускорения для задач информационной безопасности.

Общие проблемы создания реконфигурируемых средств. Одним из основных недостатков программируемой логики является высокая сложность разработки цифровых схем для ПЛИС. Данная задача включает в себя ряд трудоемких процедур. Сюда входят, в общем случае, операции создания проекта, ввода данных в специализированное инструментальное ПО, отладка, компиляция и тестирование проекта, его верификация, моделирование и тестирование, проверка работоспособности, а также оценка временных показателей созданной схемы, параметров энергопотребления и др. [10].

При создании аппаратных компонентов систем обнаружения вторжений за счет конкретизации решаемой технической задачи, этот процесс можно упростить, выполнив ряд операций зара-

нее. В итоге вся технологическая цепочка создания цифровой схемы в ПЛИС сводится к двум этапам, первый из которых – синтез вычислительной структуры – зависит от входных данных (размера и состава базы данных сигнатур), а второй – генерация файлов конфигураций – может быть выполнен автоматически посредством фирменной САПР либо специализированной программы. Разделяющим звеном между этапами выступает представление разработанной цифровой схемы на одном из языков описания аппаратуры (например, VHDL).

Синтез вычислительной структуры реконфигурируемой ССОВ является нетривиальной задачей. Как показывает проведенный обзор литературных источников, сложность состоит в многочисленности и разнообразии известных решений, в нехватке обобщения и научной проработки в данной области. Повысить эффективность и удобство применения накопленного опыта при создании распознающего устройства позволит структурирование, формальное описание и систематизация имеющейся информации, а также введение метрики для сопоставления между собой решений, основанных на различных подходах.

Генерация файла конфигурации – последовательности битов (bitstream), загружаемой в микросхему ПЛИС для придания ей требуемой функциональности, включает в себя ряд вычислительно емких процедур, таких как синтез (Synthesize), трансляция (Translate), отображение (Map), размещение и трассировка (Place & Route) и собственно формирование файла конфигурации (Bitstream Generating). Эти процедуры выполняются без участия разработчика с использованием либо фирменного пакета САПР от производителя ПЛИС, либо специального программного обеспечения.

Сложность обоих этапов – синтеза вычислительной структуры и генерации файлов конфигурации – характерна для любой области приложений реконфигурируемых вычислений.

Первый из них требует от разработчика высокой квалификации, знаний и навыков в области электроники, цифровой схемотехники, владения сложными современными программными инструментами. Второй этап требует больших вычислительных затрат на создание конфигураций для ПЛИС. В зависимости от сложности синтезируемой схемы и типа ПЛИС этот процесс может занимать от десятков минут до нескольких часов. Особенно критичной трудоемкость компиляции становится в случаях, когда задействованные ресурсы стремятся почти полностью занять площадь кристалла программируемой СБИС. В подобной ситуации процедура размещения и трассировки превращается в комбинаторно сложную задачу перебора астрономически большого числа вариантов.

Специфика РАСИБ. Сигнатурные системы информационной безопасности в качестве объектов реконфигурируемого синтеза, во-первых, отличаются высокой степенью переменчивости, обусловленной неизменно усиливающейся активностью злоумышленников и ростом их возможностей. Обновление баз данных сигнатур, вызванное появлением новых атак и вредоносных программ, производится с частотой от нескольких раз в месяц для ССОВ до нескольких раз в сутки для антивирусов. (База данных сигнатур свободно распространяемой системы обнаружения вторжений Snort, например, насчитывает несколько десятков тысяч правил). Во-вторых, процессу защиты информации в компьютерных системах присуще разнообразие многочисленных настроек и режимов работы программного обеспечения. При их изменении системный администратор оперативно включает либо выключает соответствующие записи в базе данных сигнатур ССОВ. В обоих случаях меняется состав сигнатур, распознаваемых системой защиты. Следовательно, затратную процедуру перепрограммирования реконфигурируемого устройства необходимо повторять каждый раз, как при обнаружении и описании новой атаки, так и при изменении свойств и режимов работы защищаемого объекта.

Анализ компонентов систем обнаружения вторжений и выполняемых ими функций, проведенный в п. 4.3 приводит к выводу, что при изменении исходных данных перекомпиляции подлежит только модуль распознавания, включающий в себя на аппаратном уровне базу данных сигнатур. Остальные компоненты остаются неизменными и могут быть сформированы заранее.

Современные ПЛИС обладают мощными вычислительными ресурсами (насчитывая миллионы эквивалентных логических элементов), а также высокой гибкостью. Но трудность в том, что пользователи систем информационной защиты (системные администраторы, обслуживающий персонал, лица, ответственные за информационную защиту в эксплуатирующих организациях и службах) не являются специалистами в области создания быстродействующих реконфигурируемых устройств, способными оперативно разраба-

тывать и загружать в ПЛИС нужные конфигурации.

5.2. Принципы организации процесса централизованного синтеза. Для преодоления рассмотренных выше сложностей рассматриваемый в настоящей работе вычислительный процесс организован таким образом, чтобы ресурсоемкая процедура синтеза цифровой схемы выполнялась не локально на каждой пользовательской системе информационной защиты, а централизованно – с использованием высокопроизводительных ресурсов (рис. 2).

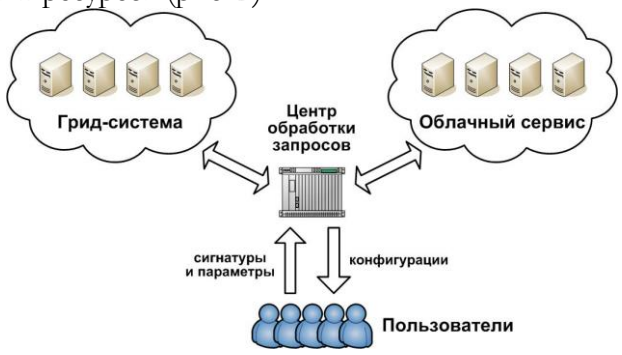

Рис. 2. Структурная схема централизованной системы синтеза РАСИБ

Единый центр обработки запросов от большого числа пользователей в этом случае выполняет следующий функции:

– оперативно пополняет базы сигнатур актуальной информацией о недавно выявленных фактах злонамеренной активности (атаках);

– собирает данные о текущих параметрах безопасности каждого из защищаемых объектов;

– осуществляет синтез цифровых схем с учетом особенностей каждой клиентской ССОВ, генерацию и оперативную доставку пользователям файлов конфигураций для загрузки в ПЛИС.

Заметим, что в качестве базиса для реализации перечисленных функций помимо грид-системы и облачного сервиса могут быть задействованы и другие высокопроизводительные платформы.

5.3. Преимущества и недостатки централизации. К преимуществам, которые предоставляет централизованный подход, следует отнести следующее.

1. За счет использования суперкомпьютерной техники повышается производительность системы в целом. Такие современные технологии как грид-системы и облачные вычисления предоставляют ресурсы, достаточные для быстрого выполнения вычислительно сложных задач оптимизации и синтеза параллельных распознающих структур на базе современных реконфигурируемых СБИС, последние семейства которых содержат миллионы эквивалентных логических элементов программируемых ресурсов, блоки внутренней памяти, аппаратные умножители и другие высокотехнологичные компоненты.

2. Благодаря разделению труда улучшаются технические характеристики локальных систем защиты информации. Централизация позволяет задействовать высококвалифицированных специалистов, результаты работы которых будут использоваться на каждой из локальных систем, что невозможно достигнуть при индивидуальной разработке систем защиты по отдельности.

3. За счет снижения требований к квалификации персонала локальных систем также снижается совокупная стоимость владения.

4. Путем группирования схожих запросов сокращаются общие вычислительные затраты. Несмотря на то, что для аппаратных ускорителей каждой клиентской системы в общем случае требуется своя собственная, уникальная конфигурация, общая база сигнатур и ограниченность числа типов используемых ПЛИС позволяют оптимизировать вычислительный процесс таким образом, чтобы многие из создаваемых комплектов реконфигурируемых компонентов могли (при незначительной избыточности) подходить одновременно нескольким клиентам. В итоге общая вычислительная сложность решения задачи снижается, причем тем существеннее, чем большее число информационных систем будет охвачено сервисом.

В качестве недостатка можно отметить усложнение системы в целом и некоторое снижение оперативности переконфигурации аппаратных компонентов, что, впрочем, в полной мере компенсируется перечисленными преимуществами.

6. Реализация системы централизованного синтеза РАСИБ на различных вычислительных платформах

6.1. Грид-сервис STRAGS. В течение нескольких последних лет при финансовой поддержке Целевой комплексной программы научных исследований НАН Украины «Грид-инфраструктура и грид-технологии для научных и научно-прикладных применений» в Институте проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины проводились работы по созданию и исследованию в среде Украинского национального грида (УНГ) сервиса централизованного синтеза конфигураций для аппаратных устройств информационной безопасности [14].

Данная разработка получила название STRAGS (Security Tasks Reconfigurable Accelerators Grid-Service – грид-сервис для реконфигурируемых ускорителей задач информационной безопасности). В качестве базиса сервис использует отечественную грид-технологию Rainbow ("ARC in the Cloud") [15], которая изначально создавалась для запуска специализированного ПО moldyngrid в интересах виртуальной организации medgrid. Разработка оказалась удачной и в последствии нашла более широкое применение в инфраструктуре УНГ. Технология позволяет запускать на грид-узлах виртуальные машины с необходимым программным обеспечением и обеспечивает к ним интерактивный доступ.

В процессе функционирования грид-сервис STRAGS в качестве агентов инициирует работу на удаленных узлах грид-среды нескольких виртуальных машин с предустановленным и настроенным инструментальным программным обеспечением, необходимым для синтеза реконфигурируемых устройств и генерации конфигураций. По мере поступления запросов от клиентов сервис распределяет задания между активными агентами, поддерживая их число достаточным для обеспечения готовности на требуемом уровне. Получив задание в виде грид-задачи, агент запускает процессы автоматического синтеза требуемой цифровой схемы и синтеза соответствующей конфигурации для ПЛИС, после чего возвращает результат работы сервису.

В работе [16] рассмотрены различные варианты программного обеспечения, используемого для синтеза цифровых схем реконфигурируемых вычислителей. В результате проведенного анализа в качестве инструментального средства было выбрано решение на основе фирменных САПР. Данный вид ПО включается в состав унифицированного образа виртуальной машины поверх системного программного обеспечения. Такой образ может быть успешно запущен на любом из грид-узлов УНГ, поддерживающем технологию Rainbow.

Важно отметить, что в процессе разработки грид-сервиса были отработаны и отлажены механизмы интерактивного взаимодействия пользователя с запущенным на виртуальных машинах инструментальным программным обеспечением. Данное качество сервиса, обеспечиваемое технологией Rainbow, позволяет пользователям в реальном времени наблюдать за ходом выполнения стадий компиляции проекта, что сводит к минимуму неудобства удаленной работы. На рис. 3 представлено одно из окон грид-сервиса, отображающее прогресс выполнения запущенного на синтез задания. Приведен момент выполнения подпроцедуры трассировки процедуры PAR (Place and Route).

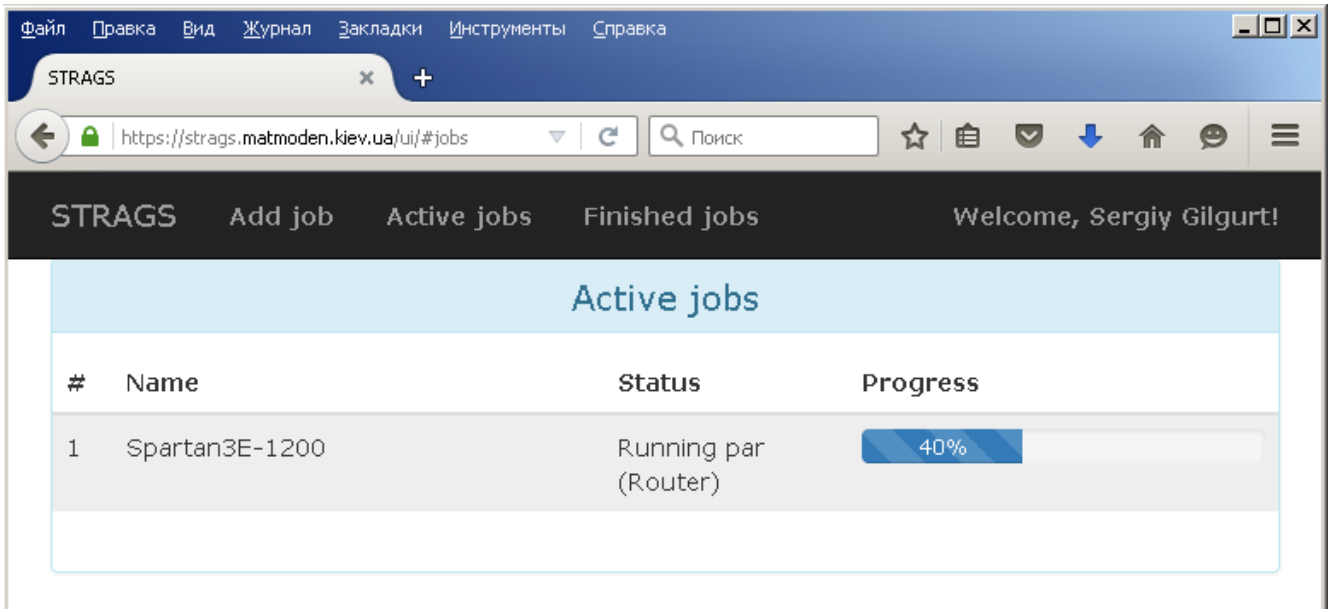

Рис. 3. Экран активных заданий грид-сервиса STRAGS

6.2. Облачный сервис. Впоследствии грид-сервис STRAGS был дополнен режимом использования технологии облачных вычислений. В качестве экспериментальной платформы был задействован сервис Elastic Compute Cloud (EC2) инфраструктуры облачных веб-сервисов Amazon Web Services (AWS) в бесплатном демонстрационном режиме.

В отличие от использования грид-ресурсов активация виртуальных машин – агентов производится не в автоматическом режиме, а по командам администратора сервиса. Состав виртуальной машины для запуска в облачной инфраструктуре доработан, во-первых, для уменьшения ее размера (с целью минимизации расходуемых в демонстрационном режиме ресурсов), во-вторых, для обеспечения взаимодействия с грид-сервисом из облачной среды. Как и для грида, виртуальные машины – агенты для облачной среды содержат инструментальное программное обеспечение, необходимое для синтеза цифровых схем и генерации загружаемых в ПЛИС конфигураций для РАСИБ.

Масштабируемая архитектура грид-сервиса STRAGS позволяет инициировать как в гриде, так и в облачной среде произвольное число агентов, ограничиваемое лишь объемом доступных вычислительных ресурсов.

6.3. Сравнение используемых технологий. Проведенные испытания сервиса STRAGS позволили провести предварительное оценочное сравнение двух высокопроизводительных технологий – грида и облачных вычислений. С поправкой на экспериментальный характер работы сервиса можно сделать следующие выводы.

Грид-среда в общем случае способна предоставить большее количество вычислительных ресурсов. В настоящее время технологию Rainbow поддерживает четыре грид-узла УНГ общей вычислительной мощностью в несколько десятков процессорных ядер. За три года проведения экспериментов в грид-среде ни разу не возникло ситуации, чтобы все четыре грид-сайта были недоступны. При этом на каждом из соответствующих кластеров простаивало (т.е. не использовалось ни под локальные задачи, ни для выполнения внешних грид-заданий) в среднем порядка 75% вычислительных ресурсов. Число активных агентов в отдельные моменты временны было доведено до нескольких десятков без каких-либо негативных последствий как для грид-сервиса STRAGS, так и для УНГ в целом. Таким образом, была косвенно подтверждена основная идея грид-вычислений, заключающаяся в возможности использовать незадействованные ресурсы. Более высокая сложность процесса организации вычислений в гриде по сравнению с облачной платформой никаким образом не сказалась на результатах тестирования опытной платформы. После модернизации грид-сервиса, проведенной через год опытной эксплуатации, система работала в круглосуточном режиме без единого сбоя в течение последующих двух лет.

Ограничения, накладываемые облачной платформой, были связаны с урезанными возможностями бесплатной ознакомительной версии коммерческого сервиса. Как следствие в процессе испытаний в облачной среде удавалось запустить не более четырех виртуальных машин одновременно. При этом к надежности данной технологии за три месяца опытной эксплуатации также не возникло никаких претензий.

Выводы

В настоящей работе исследована идея эффективного способа организации вычислительного процесса синтеза аппаратных средств информационной безопасности на базе ПЛИС. Суть решения заключается в переносе сложной и ресурсоемкой операции создания реконфигурируемых устройств с локальных систем защиты на высокопроизводительные платформы, в качестве которых задействованы грид-среда и облачный сервис.

Для решения задачи на основе анализа мирового опыта в области разработок средств киберзащиты на базе ПЛИС исследован процесс построения реконфигурируемой структуры аппаратного устройства информационной безопасности. В результате теоретического исследования сформулирована задача множественного распознавания строк. Проанализированы требования, предъявляемые к сигнатурным системам распознавания на базе ПЛИС. Приведена обобщенная структура построения системы ССОВ на РАСИБ. Выявлены преимущества и недостатки централизованного подхода.

Проведенные на разработанном прототипе системы эксперименты позволили впервые провести предварительное качественное сравнение двух задействованных высокопроизводительных платформ.

Заметим в заключение, что помимо приложений информационной безопасности аппаратное ускорение задачи множественного распознавания актуально также для таких не мене важных применений как: интеллектуальный поиск данных (data mining), ускоренная обработка XML-запросов, анализ молекул ДНК и т.п. Большой интерес данные средства представляют и для передовых сетевых приложений – классификации пакетов с восстановлением соединений для управления технологией QoS, фильтрации IP-телефонии, измерения трафика, оптимизации кеширования и не только.

ЛИТЕРАТУРА

[1]. А. Лукацкий, *Обнаружение атак*, СПб.: БХВ-Петербург, 2001, 624 с.

[2]. А. Корченко, О. Заріцький, Т. Паращук, В. Бичков, "Програмне забезпечення формування еталонів параметрів", *Захист інформації*, Т. 20, № 3, С. 133-148, 2018.

[3]. С.Я. Гильгурт, "Реконфигурируемые вычислители. Аналитический обзор", *Электронное моделирование*, Т. 35, № 4, С. 49-72, 2013.

[4]. Ю.М. Коростиль, С.Я. Гильгурт, "Принципы построения сетевых систем обнаружения вторжений на базе ПЛИС", *Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України,* Вип. 57, С. 87-94, 2010.

[5]. А.Н. Давиденко, С.Я. Гильгурт, В.И. Сабат, "Аппаратное ускорение алгоритмов сигнатурного обнаружения вторжений в открытой системе информационной безопасности Snort", *Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України*, Вип. 65, С. 94-103, 2012.

[6]. Ю.М. Коростиль, С.Я. Гильгурт, О.М. Назаренко, "Анализ базы данных системы информационной безопасности Snort и вопросы быстродействия", *Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України*, Вип. 66, С. 77-84, 2012.

[7]. B Smyth, *Computing Patterns in Strings*, Essex: Pearson Addison Wesley, 2003, 440 p.

[8]. W. Jiang, V.K. Prasanna, "Scalable Multi-Pipeline Architecture for High Performance Multi-Pattern String Matching", *24th IEEE International Parallel and Distributed Processing Symposium (IPDPS '10),* United States, Atlanta, April 19, 2010.

[9]. A.V. Aho, M.J. Corasick, "Efficient string matching: an aid to bibliographic search", *Proceedings of the II Communications of the ACM*, 1975, vol. 18, № 6, pp. 333-340.

[10]. C. Maxfield, *The Design Warrior's Guide to FPGAs: Devices, Tools and Flows,* Oxford, UK: Elsevier Science & Technology Books, 2004, 542 p.

[11]. H. Chen, Y. Chen, D.H. Summerville, "A Survey on the Application of FPGAs for Network Infrastructure Security", *IEEE Communications Surveys and Tutorials,* pp. 541-561.

[12]. А.В. Палагин, В.Н. Опанасенко, *Реконфигурируемые вычислительные системы: Основы и приложения,* К.: «Просвіта», 2006, 280 с.

[13]. С.Я. Гильгурт, "Задача множественного распознавания строк в интенсивном потоке данных и методы ее аппаратного ускорения", *Тез. доп. Міжнар. наук.-техн. конф. «Моделювання-2016»,* Київ, 2016, С. 166-169.

[14]. В.Ф. Євдокимов, А.М. Давиденко, С.Я. Гільгурт, "Створення на базі грід-сайту ІПМЕ ім. Г.Є. Пухова НАНУ системи централізованого синтезу апаратних прискорювачів для вирішення задач інформаційної безпеки в енергетичній галузі", *Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України,* Вип. 79, С. 3-8, 2017.

[15]. А.А. Сальников, В.В. Вишневский, А.Ф. Борецкий "«Платформа как сервис» в грид для интерактивного анализа медицинских данных", *Математичні машини і системи*, № 1, С. 53-64, 2015.

[16]. А.К. Гиранова, "Анализ программного обеспечения реконфигурируемых вычислителей", *Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ ім. Г.Є. Пухова НАН України,* Вип. 41, С. 43-48, 2007.

ЦЕНТРАЛІЗОВАНИЙ СИНТЕЗ РЕКОНФІГУРОВНИХ АПАРАТНИХ ЗАСОБІВ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ НА ВИСОКОПРОДУКТИВНИХ ПЛАТФОРМАХ

Основна функція сигнатурних мережевих систем виявлення вторгнень (ССОВ) – пошук в інтенсивному потоці даних ознак відомих атак з бази сигнатур, що містять останнім часом десятки тисяч записів. На відміну від міжмережевого екрану ССОВ обробляє не тільки заголовки, а й тіла пакетів, виконуючи так званий глибокий аналіз пакетів – DPI. Задача множинного розпізнавання рядків – особливий тип розпізнавання, коли вихідний текст аналізується одночасно на наявність множини патернів. У зв'язку зі стагнацією частоти мікропроцесорів, а також постійним зростанням мережевого трафіку й збільшенням кількості та складності атак традиційним програмним рішенням все складніше відповідати посилюються вимогам інформаційної безпеки. Тому все більшого поширення набувають апаратні рішення з використанням реконфігуровних пристроїв на базі ПЛІС типу FPGA, яки поєднують в собі близьку до апаратної продуктивність із гнучкістю програмного забезпечення. На жаль, розробка комплексних реконфігуровних пристроїв є нетривіальною задачею. Користувачі МСВВ, якими зазвичай є системні адміністратори, не мають для цього ні достатньої кваліфікації, ні необхідних обчислювальних ресурсів. З іншого боку, специфіка завдань інформаційної безпеки постійно вимагає виконання процедури пересинтезування реконфігуровних прискорювачів. Для вирішення проблеми було запропоновано централізувати обчислювальний процес з використанням грід-інфраструктури та хмарної платформи. Такий підхід дозволяє перенести трудомістку та обчислювально складну процедуру з локальних обчислювальних мереж до високопродуктивного середовища. Було розроблено та протестовано експериментальний сервіс. Наведено перші отримані результати. Проведено попереднє порівняння грід- та хмарної технологій. Крім кібербезпеки, прискорення задачі множинного розпізнавання актуально й для багатьох інших важливих застосувань, таких як інтелектуальний аналіз даних (data mining), прискорена обробка XML-запитів, управління технологією QoS, фільтрація в IP-телефонії, оптимізація кешування, тощо.

Ключові слова: інформаційна безпека, МСВВ, глибокий аналіз пакетів, множинне розпізнавання, ПЛІС, централізований синтез, грід, хмарні обчислення.

SYNTHESIS OF RECONFIGURABLE INFORMATION SECURITY HARDWARE ON HPC PLATFORMS

The main purpose of a signature-based network intrusion detection system (NIDS) is to inspect network packet contents against tens of thousands of predefined malicious patterns. Unlike the firewall, NIDS examines not only packet headers, but also the packet bodies. The multi-pattern string matching task is a specific type of string matching functionality to search an input stream for a set of patterns rather than a single pattern. Due to rising traffic rates, increasing number and sophistication of attacks and the collapse of Moore's law for sequential processing, traditional software solutions can no longer meet the high requirements of today's security challenges. Therefore, hardware approaches are proposed to accelerate pattern matching. Combining the flexibility of software and the near-ASIC performance, reconfigurable FPGA-based devices have become increasingly popular for this purpose. Unfortunately, the development of complex reconfigurable devices is a very difficult craft. Users of NIDS which are usually system administrators have not neither enough qualification, nor computing resources to fulfill such a work. On the other hand specificities of security tasks require frequent execution of dynamic re-synthesis of reconfigurable accelerators. To solve this problem, a centralized system based on GRID and Cloud platforms was proposed. Such approach moves design and computation complexities from LANs to HPC. An experimental system was constructed and tested. First results are received and discussed. Preliminary comparison of GRID and Cloud technologies is made. Besides cybersecurity, high-speed multi-pattern matching is required for such important applications as data mining, XML switching, QoS management, VoIP filtering, cache replication etc.

Keywords: information security, NIDS, deep packet inspection, multi-pattern string matching, FPGA, centralized synthesis, HPC, GRID, cloud computing.

Євдокимов Віктор Федорович, член-кореспондент НАН України, заслужений діяч науки и техніки України, доктор технічних наук, професор, почесний директор Інституту проблем моделювання ім. Г.Є. Пухова НАН України, головний науковий співробітник відділу Математичного і комп'ютерного моделювання Інституту проблем моделювання ім. Г.Є. Пухова НАН України.
E-mail: evdokimovvf@ipme.kiev.ua.

Евдокимов Виктор Федорович, член-корреспондент НАН Украины, заслуженный деятель науки и техники Украины, доктор технических наук, профессор, почетный директор Института проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины, главный научный сотрудник отдела Математического и компьютерного моделирования Института проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины.

Evdokimov Viktor, Corresponding Member of the National Academy of Sciences of Ukraine, Honored Worker of Science and Technology of Ukraine, Doctor of Technical Sciences, Professor, Honorary Director of the Pukhov Institute for Modelling in Energy Engineering of NAS of Ukraine, Chief Researcher of the Department of Mathematical and Computer Modelling, Pukhov Institute for Modelling in Energy Engineering of NAS of Ukraine.

Давиденко Анатолій Миколайович, кандидат технічних наук, старший науковий співробітник, провідний науковий співробітник відділу Теорії моделювання Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України.
E-mail: davydenko@ipme.kiev.ua.

Давиденко Анатолий Николаевич, кандидат технических наук, старший научный сотрудник, ведущий научный сотрудник отдела Теории моделирования Института проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины.

Davydenko Anatoly, Candidate of Technical Sciences, Senior Researcher, Leading Researcher of Department of Modelling Theory, Pukhov Institute for Modelling in Energy Engineering of NAS of Ukraine.

Гільгурт Сергій Якович, кандидат технічних наук, старший науковий співробітник, старший науковий співробітник відділу Теорії моделювання Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України.
E-mail: hilgurt@ipme.kiev.ua.

Гильгурт Сергей Яковлевич, кандидат технических наук, старший научный сотрудник, старший научный сотрудник отдела Теории моделирования Института проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины.

Hilgurt Serhiy, Candidate of Technical Sciences, Senior Researcher, Senior Researcher of Department of Modelling Theory, Pukhov Institute for Modelling in Energy Engineering of NAS of Ukraine.