DOI 10.36622/VSTU.2022.25.1.005
УДК 004.056.57

ПЕРСПЕКТИВЫ ПОСТРОЕНИЯ ГЕНЕРАЛЬНОЙ КАРТЫ ЗАЩИЩАЕМОГО КИБЕРПРОСТРАНСТВА

А.Л. Сердечный

В статье рассматривается представление генеральной карты киберпространства на его физическом, информационном и социальном уровнях. Применительно к физическому уровню обозначены ландшафты операторов, зон взаимодействия устройств интернета вещей, расположения киберфизических систем. В номенклатуру ландшафтов информационного уровня включены: данные о логических маршрутах с учетом различных сетевых протоколов; альтернативные протоколы, организующие работу обособленных виртуальных пространств; карты содержимого информационных ресурсов, онтологий и банков данных. На социальном уровне отмечается обилие объектов с множественными межслойными пересечениями. Рассматриваются также координаты объектов киберпространства через идентификаторы уровня, типового слоя, набора данных, алгоритма проецирования слоя, изображения объекта в рамках слоя. Слои кибепространства предлагается определять с учетом возможности масштабирования. Проекции слоев киберпространства позволяют рассматривать одни и те же объекты в различных вариантах расположения на плоскости или в трехмерном пространстве. Масштабирование слоев киберпространства позволяет организовывать направленную работу с большими данными. Вниманию читателя предлагаются многочисленные примеры, иллюстрирующие вышеизложенные методические рекомендации.

Ключевые слова: генеральная карта, уровни, идентификаторы, слои, объекты, масштабирование, проекции.

Введение

Появление проекции Меркатора позволило картографам договориться об универсальном языке изображения географических объектов [1, 2]. В результате географическая карта превратилась в удобный инструмент, позволяющий планировать сложные экспедиции в различные точки планеты.

Атрибутами географической карты являются [3, 54]:

- ландшафт земной поверхности, отображаемой на географической карте;
- сведения о проекции, в которой изображена земная поверхность (для того, чтобы иметь представления об искажениях, обусловленных переходом из трёхмерного в двухмерное пространство и иметь возможность измерения необходимых параметров на основании карты);
- масштаб и координатная сетка, позволяющие позиционировать объекты географического пространства и определять расстояния между ними;
- условные обозначения, позволяющие передать дополнительные сведения об изображаемых на карте географических объектов;
- сведения об авторе (издательстве) и другая дополнительная информация, позволяющая определить степень доверия к данным, изображаемым на географической карте.

Для киберпространства требуется свой язык и генеральная карта, которая бы объединила различные проекции информационных карт. Вместо ландшафта земной поверхности в такой карте выступают ландшафты слоёв киберпространства. Каждый слой отражает однотипные взаимосвязанные объекты на одном из уровней киберпространства. Эти объекты образуют ландшафт слоя (рис. 1).

Необходимо рассматривать не менее трёх групп взаимосвязанных уровней киберпространства: физический, информационный и социальный.

На физических уровнях могут быть отражены следующие ландшафты:

- ландшафт операторов связи, предоставляющих доступ к сети Интернет (с привязкой к географической карте и в виде графа физических линий связи);

- ландшафт зон взаимодействия IoT-устройств (устройства, реализующие технологию Интернета вещей);
- ландшафт расположения киберфизических систем.

В настоящий момент традиционным подходом картографирования сети Интернет на физическом уровне является изображение инфраструктуры интернет-провайдеров. Развитие беспроводных технологий и средств автоматизации привело к трансформации услуг глобальной связи.

Всё большей популярностью начинают пользоваться технологии децентрализованного взаимодействия устройств при помощи встроенных беспроводных модулей. Автоматизация транспортных средств и других ранее нецифровизованных объектов привела к появлению киберфизических систем. Такие объекты построены с использованием большого количества технологий и представляют собой многокомпонентные информационные системы. Например, судно с киберподдержкой может включать интегрированную мостиковую систему, автономную навигационную систему, систему слежения за контейнерами и другие компоненты, размещённые по всему кораблю и связанные между собой с помощью десятка различных протоколов.

Номенклатура ландшафтов информационных уровней должна быть существенно шире физических. Наиболее низкие уровни (имеющие наиболее тесные связи с физическими уровнями) должны представлять информацию о логических маршрутах с учётом различных сетевых протоколов. Для сети Интернет основными протоколами являются [5]:

- BGP (маршрутизация трафика между автономными системами – информационно-телекоммуникационными системами Интернет-провайдеров);
- IP4 и IP6 (маршрутизация трафика между устройствами);
- DNS (идентификация информационных ресурсов).

Для локальных вычислительных сетей, а также устройств интернета вещей и конкретных киберфизических систем также имеет смысл рассмотривать протоколы ARP и ICMPv6, позволяющие строить графы связей и взаимодействий узлов сети на основании аппаратного адреса (MAC-адреса) [6]. Однако с появлением и активным внедрением технологии рандомизации (случайной смены) MAC-адресов эффективность использования соответствующих карт постепенно снижается.

Кроме того, на информационном уровне должны учитываться альтернативные сетевые протоколы, которые с использованием базовых Интернет-технологий организуют работу обособленных виртуальных пространств. К таким технологиям относятся [6]:

- BitTorrent (децентрализованный протокол обмена файлами);
- The Onion Router (технология «луковой» маршрутизации для организации анонимных сетей);
- I2P (стек протоколов «чесночной» маршрутизации для организации децентрализованной анонимной сети) и др.

Также к информационному уровню относятся информационные карты содержимого информационных ресурсов, онтологий и банков данных. Одним из важных слоёв данного уровня является слой тематик научных публикаций, которые отражают все аспекты исследованию защищаемого киберпространства.

На социальных уровнях количество взаимосвязанных объектов ещё шире, так как появляется прикладной аспект использования киберпространства. Каждая социальная платформа предоставляет свои уникальные возможности и содержит разнообразный контент, обладающий специфическими свойствами. Между слоями социальных уровней могут быть множественные пересечения, которые позволяют взглянуть на объект исследования под новым углом и выявить новые закономерности. Так, например, изображение на кате музыкальных предпочтений участников политических сообществ позволяет выявить направленность организаторов подобных сообществ в отношении определённых категорий граждан, что затруднительно сделать на основании лишь табличных сведений.

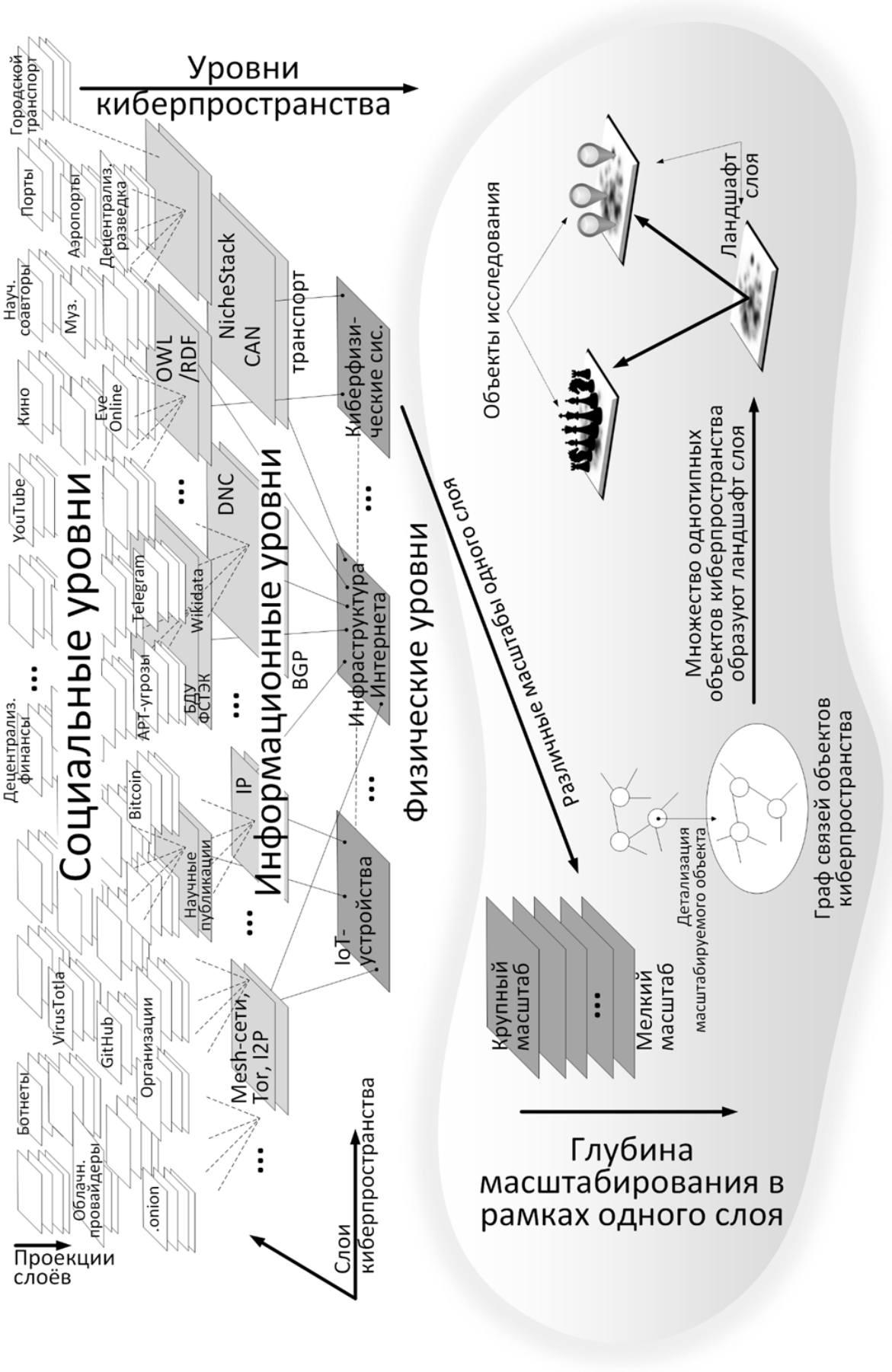

Рис. 1. Представление генеральной карты киберпространства как системы слоёв различных уровней, проекций и масштабов

Особенности разработки генеральной карты киберпространства как объекта атаки

Каждый слой карты может быть представлен в различных проекциях. Проекции строятся на основании одних и тех же исходных данных в соответствии с одинаковой моделью данных. Их отличие заключается в выборе метрики близости и алгоритма снижения размерности. К слою могут быть применены операции масштабирования, если имеется возможность детализации ландшафтообразующих объектов киберпространства.

Координаты объектов киберпространства

Одной из важных задач генеральной карты является идентификация объектов киберпространства. В качестве аналога координатной сетки (позволяющей локализовать объекты географического пространства) может быть использована следующая система идентификации:

- идентификатор уровня киберпространства (кортеж идентификаторов, если объект пересекает несколько уровней);
- идентификатор типового слоя киберпространства;
- идентификатор набора данных, на основании которого сформирован ландшафт слоя;
- идентификатор алгоритма проецирования слоя (алгоритма преобразования набора данных в ландшафт слоя) и значения входных параметров;
- координаты изображения объекта киберпространства в рамках системы координат слоя информационной карты.

На основании данного набора признаков положение объекта киберпространства может быть однозначно установлено. Такой подход (при наличии соответствующих информационных карт и средств автоматизации) позволяет эксперту проводить анализ больших данных не только благодаря реализации алгоритмов машинного обучения, но и с учётом своего опыта и знаний.

Первоочередной задачей построения генеральной карты защищаемого киберпространства является формирование базы уровней, типовых слоёв, наборов данных и алгоритмов проецирования объектов киберпространства. Такая задача может быть решена лишь при наличии достаточного количества различных вариантов информационных карт, построенных с использованием методов и средств, описанных в настоящей работе. Некоторые примеры карт слоёв представлены ниже.

Слои киберпространства

В условиях отсутствия необходимых для решения задачи построения генеральной карты набора исходных данных, в рамках настоящей работы решено ограничится формулированием требований к составу слоёв и их ландшафтов, которые должны способствовать появлению необходимых карт. Данные требования учитывались при разработке информационных карт в рамках решения частных задач в области защиты информации и обеспечения информационной безопасности.

Деление объектов на слои должно определяться структурой их связей с другими объектами. Не имеет смысла объединять в один слой несвязанные объекты, точно также как объединение полностью связанных объектов не даст никакой выгоды от их изображения в рамках единого ландшафта.

Другим требованием к процедуре выбора слоёв является обеспечение возможности масштабирования информационных карт. Для этого состав и порядок следования слоёв должен быть подобран таким образом, чтобы взаимное расположение между изображениями одних и тех же объектов для различных слоёв не имело существенных искажений. Тогда при переходе от одного слоя к другому исследователю не придётся тратить усилия на восприятие нового контекста и поиск объектов в новых позициях.

Также важное значение при образовании ландшафта слоя имеет количество типов связей между объектами киберпространства. Если типов связей слишком много, то при визуальном анализе такого графа трудно установить причину близкого расположения узлов, так как не ясно, какие конкретно факторы (связи) определили позиции вершины. Несколько типов связей имеет

смысл использовать, когда требуется оценить совокупное влияние всех факторов. Таким образом, в большинстве случаев на одном слое должны находиться однотипные объекты, а количество различных типов должно определяться на основании локальных значений функции стресса.

Так, например, если на одном слое показать связи между соавторами научных публикаций, их принадлежностью к организациям, а также связи цитирования и ключевые слова, то получится перегруженное изображение. Визуальный анализ такого изображения не даст положительных результатов. Поэтому граф необходимо разделить на несколько частей:

- граф творческих коллективов (отражает связи между соавторами);
- граф цитирования (демонстрирует развитие научных идей);
- граф ключевых слов (позволяет установить тематическое сходство между публикациями).

В этом случае появляется возможность анализа связей между слоями. Например, если в качестве ландшафта использовать граф цитирования научных публикаций, то можно изучить тематическую направленность авторских коллективов.

Говоря о влиянии факторов, необходимо отметить, что для разных типов информационных объектов их сила может существенно отличаться. В рамках построения генеральной карты наиболее значимыми факторами являются:

1) на социальном уровне:
- культурные особенности субъектов киберпространства;
- языковые особенности субъектов киберпространства;
- индивидуальные особенности субъектов киберпространства;

2) на информационном уровне:
- семантика информационных объектов киберпространства;
- способ представления и передачи информации;

3) на физическом уровне:
- способ кодирования данных;
- пространственное расположение объектов киберпространства.

Чем ниже уровень слоя, тем выше влияние соответствующего фактора. Продемонстрируем это правило на следующем примере (рис. 2).

На рис. 2 показаны четыре карты объектов одинакового масштаба:
- карта автономных систем (рис. 2 А);
- карта музыкальных предпочтений (рис. 2 Б)
- карта поиска научных публикаций (рис. 2 В)
- карта организаций (рис. 2 Г).

Для построения карт были использованы одинаковые алгоритмы. Формирование каркаса ландшафта осуществлялось с помощью алгоритма силовой укладки графа ForceAtlas2 [8]. После укладки графа строилась тепловая карта и осуществлялась экспертная разметка областей, выявленных в результате автоматической кластеризации соответствующих графов при помощи Лейденского алгоритма [9]. Расположение кластеров, которые на рис. 2 обозначены текстом, определяется совокупностью указанных выше факторов. Эти факторы заложены в связях графов объектов.

Для карты автономных систем в качестве объектов рассматривались идентификаторы автономных систем Интернет-провайдером, а связями выступали правила маршрутизации интернет-трафика (рис. 2 А). На первый взгляд, для данной карты наибольшую силу должно иметь географическое положение инфраструктуры оператора связи. Однако на практике ключевую роль играют политические факторы, обусловленные культурными особенностями. Китай, который ввиду своего большого влияния и территории, должен иметь куда большие размеры на информационной карте. Однако при помощи «великого китайского файрвола» Китай ограничил доступ к своему информационно-телекоммуникационному пространству. Противоположную позицию занимают США, которые стремятся как можно сильнее распространить своё влияние. В результате на информационной карте можно наблюдать плотный кластер «стран Запада», большую часть которого занимает США, несмотря на физическое расстояния между государствами, входящими в данный кластер.

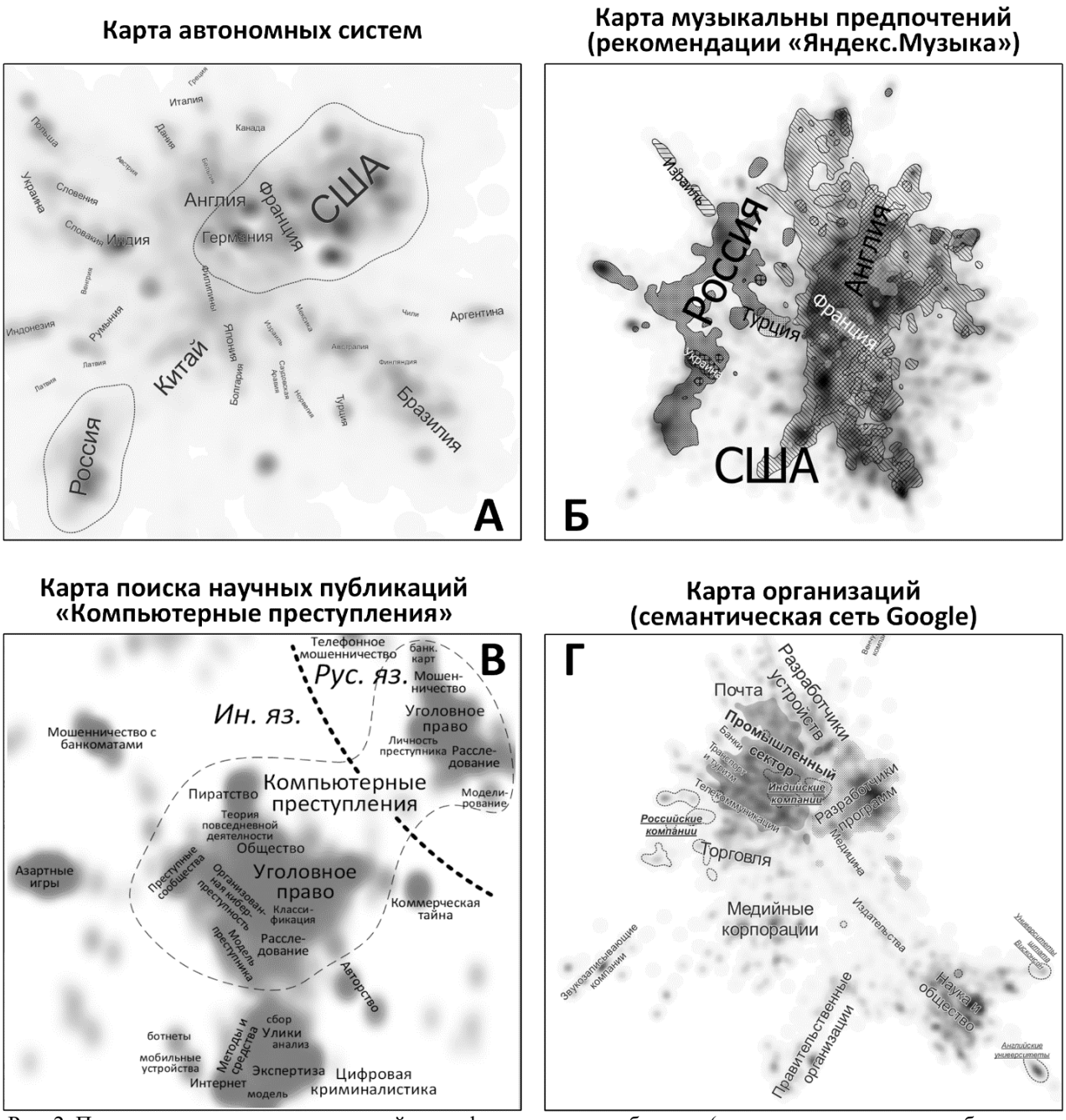

Рис. 2. Примеры влияния различных свойств информационных объектов (таких как культурные особенности, принадлежность государству, географическое положение) на их расположение в информационных картах (А – карте автономных систем, Б – карте музыкальных предпочтений, В – карте поиска научных публикаций по дисциплине «Компьютерные преступления», Г – карте организаций)

Схожую ситуацию можно наблюдать на карте музыкальных предпочтений, построенную на основании сведений информационного сервиса «Яндекс.Музыка», (рис. 2 Б).

Узлами графа являются музыкальные исполнители, а связи определяются на основании алгоритмов работы рекомендательной системы Яндекс (связи определяются на основании музыкальных вкусов пользователей и позволяют подобрать похожих артистов). Страны Запада также представлены в виде плотной группы, однако, в первую очередь это обусловлено языковыми и культурными особенностями. Так как аудиторией сервиса «Яндекс.Музыка» в основном являются представители стран СНГ, можно видеть чёткое разделение на две крупные зоны, «мостом» между которыми выступает кластер с турецкими музыкальными исполнителями. Представительство китайских исполнителей на данной карте не позволяет организовать сколько-нибудь

заметный кластер, что обусловлено сильными культурными различиями.

Также языковой фактор является ключевым при разделении на крупные зоны информационной карты научных публикаций (рис. 2 В). При этом необходимо отметить, что семантический фактор для русскоязычных и англоязычных публикаций достаточно схож, что отражается во внутренней структуре соответствующих зон.

Информационная карта, изображённая на рис. 2 Г, демонстрирует взаимосвязь между наиболее крупными компаниями. Для построения данной карты были использованы сведения из базы знаний Google, которые можно получить в результате выполнения поисковых запросов по названию организации. Компания Google не раскрывает сведений как о механизмах построения базы знаний, так и об используемых на её основе рекомендательных алгоритмов выдачи схожих организаций. Однако изучая особенности ландшафта карты можно получить некоторые полезные сведения. Так, например, на карте чётко выражен дисбаланс влияния различных факторов, что, скорее всего, обусловлено различной силой связей между объектами графа знаний Google. Так, например, есть чёткое разделение на научные, коммерческие и некоммерческие организации. В рамках коммерческого кластера идёт разделение по отраслям, что позволяет говорить о решающей роли фактора вида деятельности организаций. При этом в рамках научного кластера организации группируются в основном по географическому признаку. Влияние этого фактора также объясняет обособленность кластера «Российские компании», но уже в другой зоне карты. Однако внутри российского кластера прослеживается разделение пространства карты по виду деятельности. Способы учёта различных факторов могут существенным образом повлиять на оценки рекомендательных систем и различных рейтинги. Данную особенность наглядно демонстрируют информационные карты одного из важнейших слоёв киберпространства – слоя информационных ресурсов сети Интернет. Рассмотрим этот пример на примере двух проекций данного слоя.

Проекции слоёв киберпространства

Проекции слоёв киберпространства позволяют рассмотреть одни и те же объекты в различных вариантах расположения на плоскости или в трёхмерном пространстве. За счёт корректировки метрики близости возможно менять баланс сил между факторами, влияющими на расположение объекта. Подбор параметров алгоритма снижения размерности позволяет «увидеть» искажения, вызванные потерей информации при переходе в пространство более низкой размерности. Данные операции позволяют определить одну или несколько лучших проекций, позволяющих адекватно изобразить соответствующие информационные объекты Таким образом решается *прямая задача* – построение адекватной проекции, отражающей баланс факторов между объектами киберпространства.

Также имеет место и *обратная задача*. Зная лишь значения расстояний между объектами (значения метрики близости), но не обладая информацией о формуле её расчёта и исходных данных, можно выдвинуть гипотезу об этих параметрах на основании анализа различных проекций изображаемых информационных объектов. слоя. В результате наблюдения за различными проекциями одних и тех же объектов можно установить наиболее полный состав и наиболее точные значения всех действующих факторов.

В качестве примера сравним две информационные карты взаимосвязей наиболее популярных Интернет-ресурсов. Карты строились на одних и тех же исходных данных (графа связей информационных ресурсов сети Интернет для 5000 наиболее популярных доменных имён [12]), но отличались метриками близости. Такие информационные карты строятся на основании данных, полученных от систем мониторинга Интернет-пространства [10-18]:

- Alexa Top 1m [12] (компания Alexa, ныне подразделение Amazon, одна из первых начала составлять рейтинги Интернет-ресурсов и ранее публиковала рейтинги миллиона наиболее популярных сайтов);

- Cisco Umbrella (список миллиона самых популярных сайтов, периодически

составляемый компанией Cisco на основании сведений трафика DNS) [13];

- Majestic Million (Majestic ежедневно публикует список, составленный по результатам обхода Интернет-сайтов, который ранжируется на основании обратных ссылок) [14];

- DomCorp (компания Commonsearch на основании данных CommandCrawl [16] составила список 10 миллионов лучших сайтов) [18].

Как было показано в [10], системы мониторинга используют собственные подходы для расчёта рейтинга, причём эти подходы существенно отличаются друг от друга, о чём свидетельствуют данные взаимного пересечения соответствующих перечней (рис. 3).

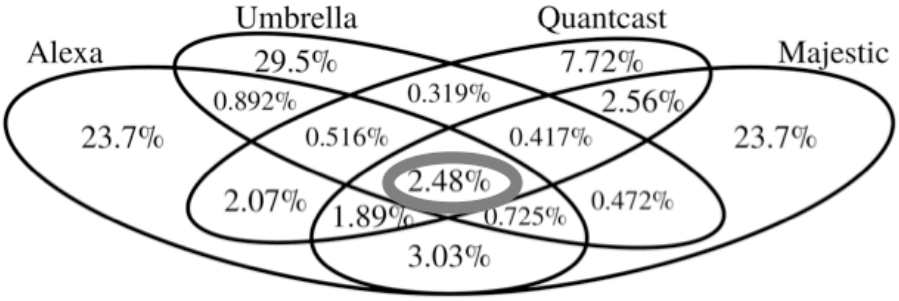

Рис. 3. Пересечение перечней наиболее популярных Интернет-ресурсов, составленных на основе данных различных систем мониторинга [10]

Построение информационных карт связей доменных имён информационных ресурсов позволяет увидеть структуру этих различий. Для построения были использованы данные рекомендательных систем, определяющих сходство сайтов [12, 14, 17]. Узлами графа являлись DNS-имя сайтов, а связи определялась при помощи рекомендательных систем (рис. 4). При этом необходимо отметить, что методика расчёта схожих сайтов не была известна, а вес связи между схожими информационными ресурсами считался равным 1. Так как рекомендательные системы использовали собственные методики расчёта сходства информационных ресурсов, то метрики близости отличались, что обеспечило различие проекций построенных информационных карт.

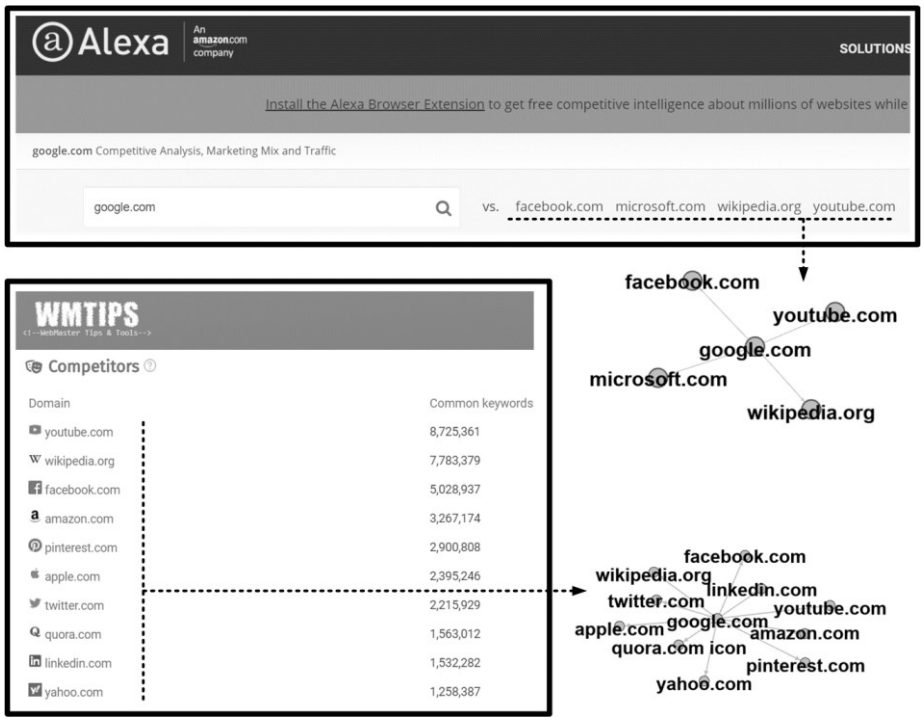

Рис. 4. Исходные данные о связях ресурсов сети Интернет и соответствующие графы

Как было установлено в ходе построения и анализа информационных карт, источники Alexa [12] и Majestic [14] в большей степени используют сходство содержимого сайтов, так как на соответствующих информационных картах образовывались кластеры по языковому признаку, причём для русского языка сходство между сайтами было определено недостаточно качественно.

В отличии от Alexa и Majestic, для информационных карт на базе MegaIndex [17] в большей степени был проявлен фактор тематического сходства кластеров. Сайты на разных языках, но со схожей тематикой располагались ближе друг к другу, чем к информационным ресурсам с идентичными языками, но разными тематиками. Информационная карта на основе тематического сходства более предпочтительна для задач исследования защищаемого киберпространства, связанного с информационным контентом. Карта на основе языкового сходства в большей степени применима к изучению влияния в киберпространстве стран и культур (рис. 5).

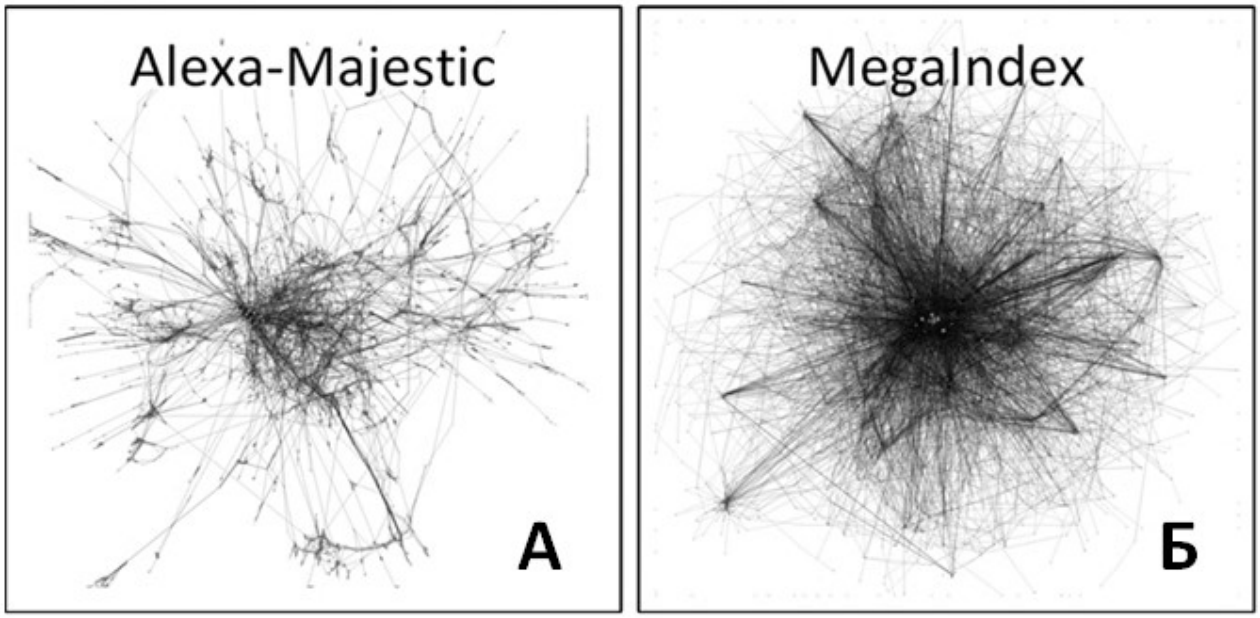

Рис. 5. Проекции информационных карт связей интернет-сайтов
(А – на основе данных Alexa [12] и Majestic [14], Б – на основе данных MegaIndex [17])

Проекции карт связей интернет-сайтов отличаются следующим. Для проекции «Alexa-Majestic» (рис. 5.А) в качестве метрики близости использовались результаты соответствующих рекомендательных систем, которые для одного сайта позволяли определить 4 похожих информационных ресурса. Для укладки графа использовался алгоритм ForceAtlas2 (theta=1.2, «устойчивость»=1, «разрежённость»=1, «гравитация»=1). Наиболее популярные сайты сконцентрированы в центре графа. Для проекции «MegaIndex» мера близости могла основываться на более чем 4 схожих сайтах. Поэтому плотность графа была больше, чем в первом случае, и для более равномерного распределения узлов в пространстве был использован режим «LinLog» силового алгоритма ForceAtlas2 (LinLog=true, theta=1.2, «устойчивость»=1, «разрежённость»=1, «гравитация»=1). Это привело к увеличению силы отталкивания в центральной области карты. Наиболее популярные сайты также оказались размещёнными в центре карты, при этом их расположение было более компактным.

На рис. 6 показана карта «Alexa-Majestic» с нанесённой на неё разметкой стран. Разметка осуществлялась автоматически на основании собранных сведений о соответствующих информационных сайтов.

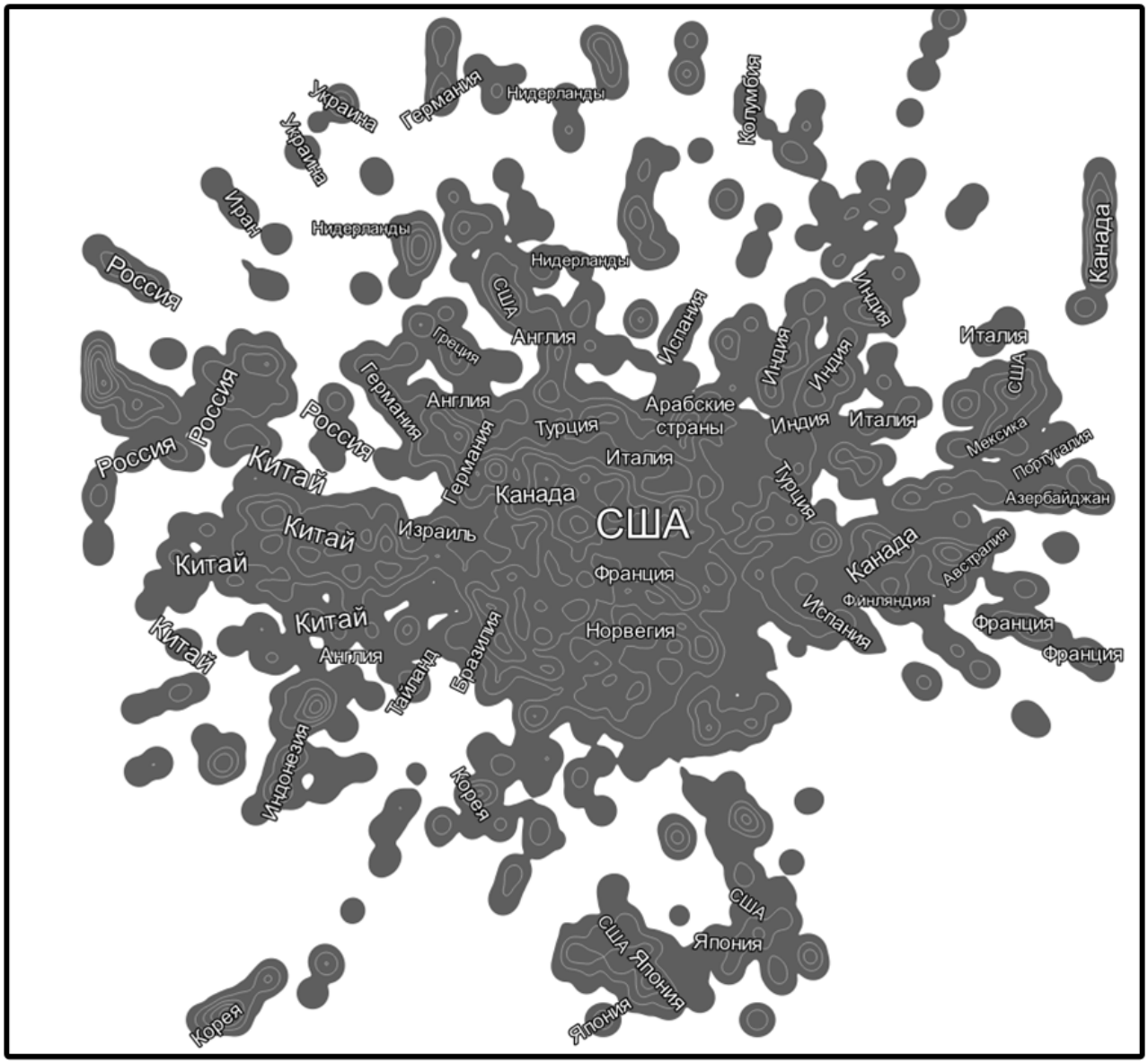

Рис. 6. Карта связей Интернет-сайтов в проекции «Alexa-Majestic» с нанесённой разметкой кластеров

Наиболее крупная зона относится к Интернет-сайтам США, которая включает в себя не только центральную часть, но и фрагменты национальных кластеров (особенно Японского). Достаточно велики кластеры России, Китая и Европы. Нет ярко выраженного кластера Англии и Франции, которые локализованы в различных частях карты.

Внутри крупных национальных кластеров наблюдается объединение узлов в соответствии с тематическим сходством. Так, например, для кластера «Россия» чётко прослеживаются тематические кластеры «Кино и сериалы», «Погода», «Спорт», «Торговля», «Финансы», «СМИ», «Госуслуги», «Онлайн-магазины» и «Работа» (рис. 7). Связь отечественных Интернет-ресурсов с ядром карты осуществляется через такие информационные ресурсы как Яндекс, телеканал RT, социальные платформы ВКонтакте и Одноклассники.

Рис. 7. Карта связей отечественных интернет-сайтов в проекции «Alexa-Majestic»

Также необходимо отметить, что из-за ошибок, вызванных совмещением размерности, существуют участки, в которых несвязанные кластеры накладываются друг на друга. Пример такого эффекта проиллюстрирован на рис. 7.

На рис. 8 показана другая проекция (карта «MegaIndex») с нанесённой на неё разметкой. Компактное расположение тематических кластеров подтверждает гипотезу о том, что тематический фактор в данном случае преобладает над языковым.

Ядро карты, расположенное в центре, включает кластеры популярных социальных платформ, средств массовой информации и информационных ресурсов, связанных с творчеством и здоровьем. Для этой области значение функции стресса имеет максимальные значение и трудно говорить о влиянии того или иного фактора.

На информационной карте «MegaIndex» имеется явно выраженное направление смены тематики. Переход из нижней части карты в правый верхний угол определяет переход тем, связанных с досугом, к теме работы и финансов. В верхней левой части расположены кластеры, связанные с IT-технологиями и разработкой программного обеспечения.

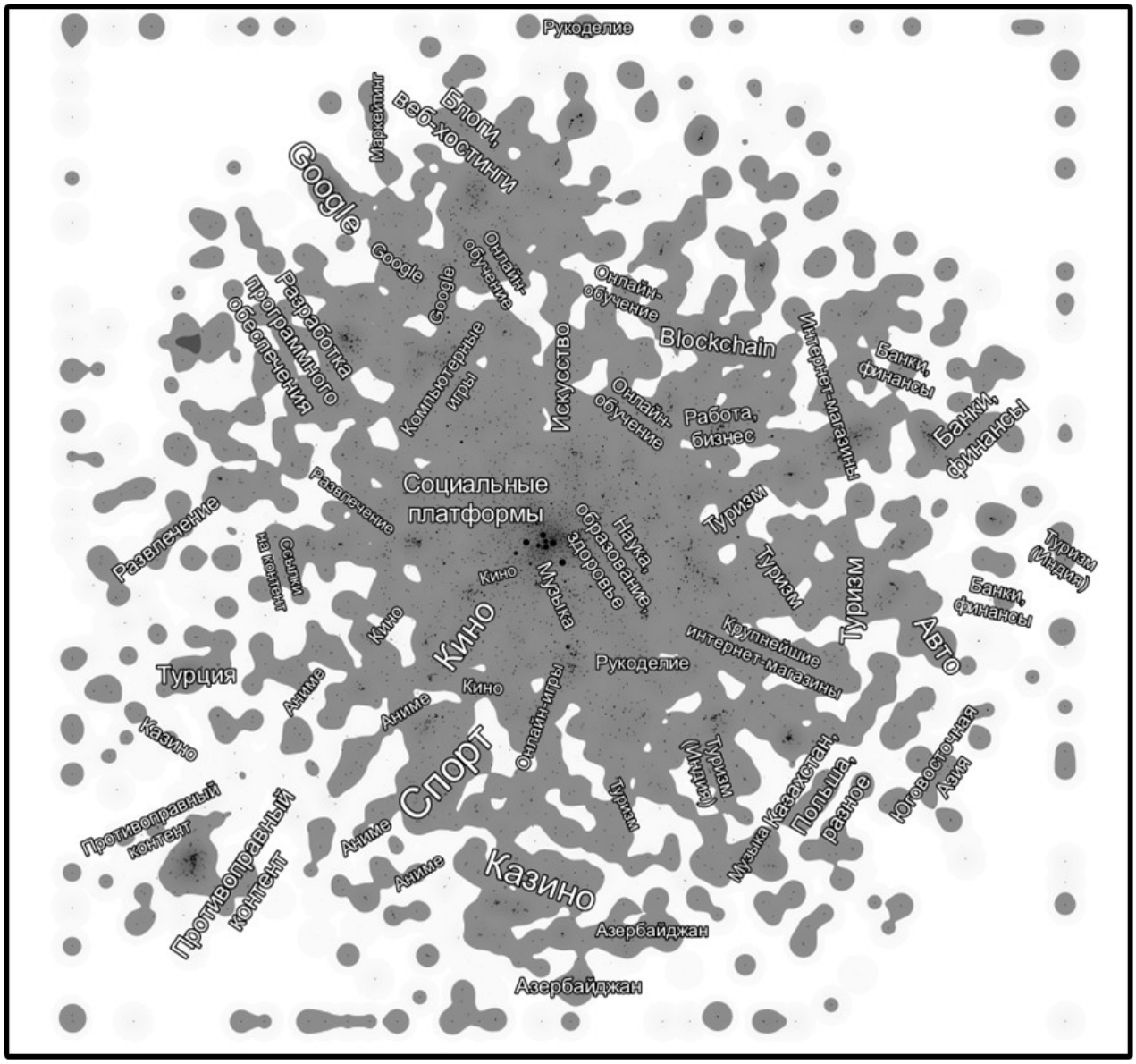

Рис. 8. Карта связей интернет-сайтов в проекции «MegaIndex» с нанесённой разметкой кластеров

Пример с системами мониторинга интернет-ресурсов показывает необходимость учёта специфики источников исходных данных при определении состава и содержания слоёв генеральной карты защищаемого киберпространства. Кроме того, важно рассматривать различные проекции, которые каждая из которых по-своему отражает связи между информационными объектами киберпространства.

Масштабирование слоёв киберпространства

Масштабирование слоёв киберпространства позволяет организовать направленную работу с большими данными. Переход между слоями и проекциями даёт возможность загружать ограниченный набор данных, необходимых для построения текущей проекции. Если в ходе её анализа возникнет потребность изучения конкретного объекта в другом слое, то возможность масштабирования позволяет подобрать нужные информационные карты с новым контекстом, причём их отображение для исследователя должно быть органично встроено в процесс исследования. То есть перемещение между слоями и проекциями защищаемого киберпространства должно быть организовано подобно тому, как в ходе масштабирования традиционных карт не нарушается целостность процесса исследований географических объектов.

Однако, если для географических карт такое масштабирование возможно только в одном направлении – от карт мелкого

масштаба к более крупным и обратно, то для киберпространства перемещение может осуществляться между его уровнями. Для демонстрации используем рассмотренный выше пример – информационные карты Интернет-связей. Дополнительно рассмотрим сведения об операторах связи, предоставляющих услуги хостинга сайтов.

Доступ к самым крупным сайтам обеспечивают одни из ведущих в мире провайдеров, таких как Cloudflare, Amazon, Hetzner, Ростелеком и др. Каждый провайдер для предоставления услуг связи имеет один или несколько идентификаторов автономных систем, с помощью которых обеспечивает маршрутизацию трафика на уровне провайдеров. Между автономными системами распределены диапазоны IP-адресов, с помощью которых осуществляется маршрутизация трафика на уровне узлов сети Интернет. В свою очередь, для взаимодействия пользователей с информационными ресурсами имеются DNS-идентификаторы (доменные имена сайтов). Как правило, операторы связи предоставляют услуги по размещению Интернет-сайтов, обеспечивая при этом требуемое качество доступа и защиту от различных угроз (DDoS-атаки и др.). Таким образом, когда пользователь с помощью браузера обращается к сайту, используя доменное имя, инициируется сложный процесс, который включает множество этапов, среди которых:

- обращение браузера пользователя к серверу доменных имён (DNS-серверу) с целью определения IP-адреса сайта;
- отправка своему провайдеру сетевых пакетов, предназначенных для доставки адресату (веб-серверу, где размещён сайт) с установлением IP-адреса;
- передача провайдером (на основании таблицы маршрутизации) сетевых пакетов пользователя в другую автономную систему (с целью доставки их до адресата наиболее быстрым способом).

Некоторые задачи защиты киберпространства требуют многогранное исследование подобных процессов. Для этого могут быть использованы масштабируемые информационные карты.

Так, например, рассмотренная выше цепочка определяет модель данных со следующими типами связей и информационных объектов:

- связь между провайдерами на основании логических маршрутов передачи пакетов (граф связей идентификаторов автономных систем между собой, а также с названиями провайдеров);
- связь между провайдерами и диапазонами IP-адресов, которые им предоставлены (граф связей идентификаторов автономных систем и IP-адресов);
- связь между провайдером и хостингом веб-сайтов, которому провайдер предоставил IP-адрес (граф связей идентификаторов автономных систем и IP-адресов хостингов веб-сайтов);
- связь между хостингом и размещаемыми сайтами (граф связей IP-адресов хостингов и доменных имён сайтов).

На основании этих связей можно формировать дополнительные комбинации. Так цепочку отношений «Имя провайдера» → «IP-адрес» → «DNS-имя» можно сократить до «Имя провайдера» → «DNS-имя» путём исключения промежуточных элементов. Но, если построить информационную карту на основе такой комбинации и использовать её для анализа связей между хостинг-провайдерами и информационными ресурсами (например, для выявления поставщиков деструктивного контента или определения области влияния хостинг-провайдера), то в некоторых случаях можно упустить из вида важные факты. Чем больше различных типов информационных объектов и чем сложнее цепочки связей между ними, тем выше вероятность допущения подобных ошибок. Гораздо удобнее использовать масштабируемые информационные карты.

Рассмотрим работу одного из крупнейших хостинг-провайдеров – компанию Cloudflare. Как видно из рис. 9, данный оператор связи предоставляет услуги большому количеству наиболее популярных сайтов. При этом есть области, в которых данный провайдер представлен в меньшей степени (информационные ресурсы Россия и Китая).

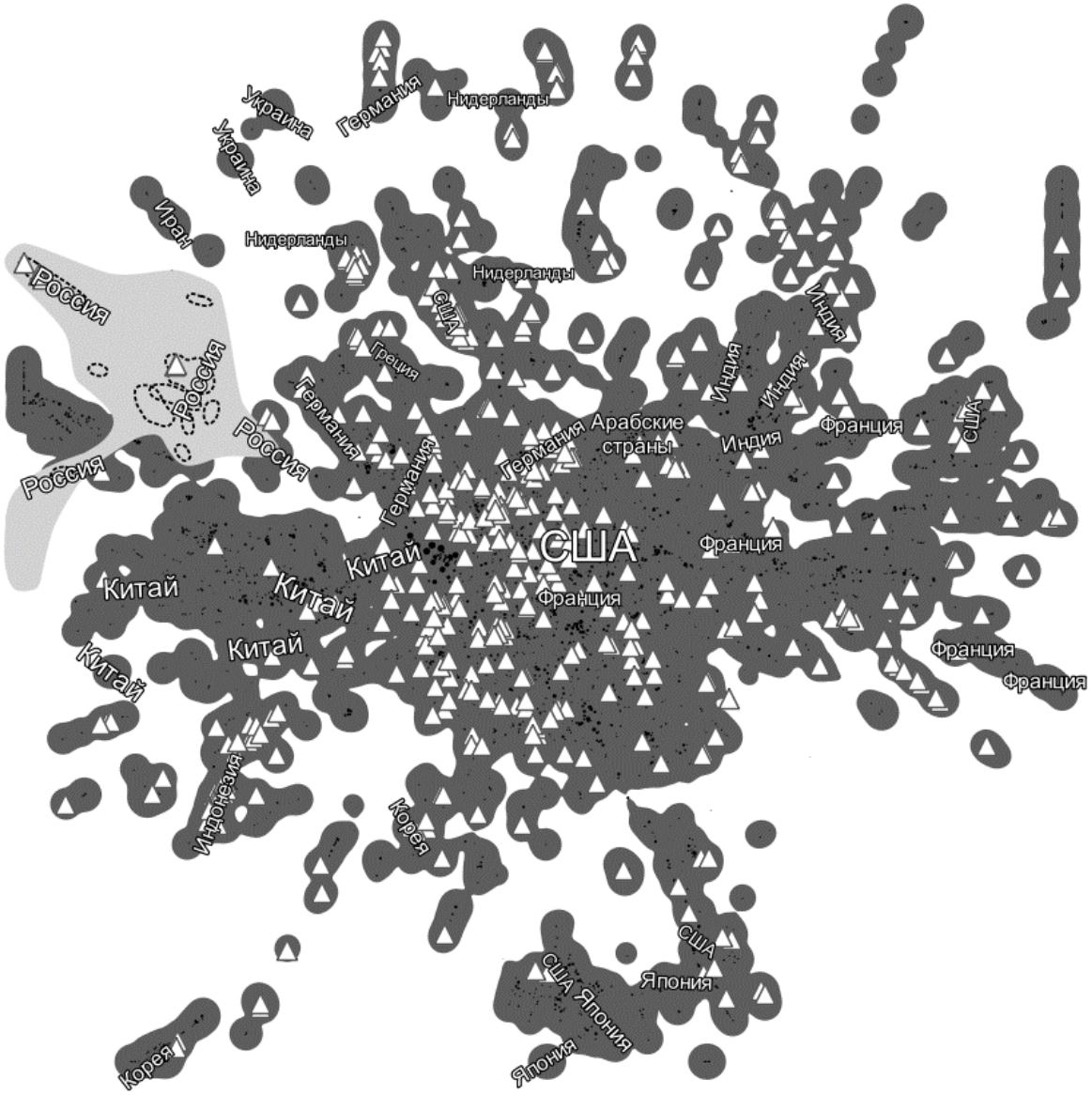

Рис. 9. Изображение (в виде треугольников) информационных ресурсов, для которых хостинг-провайдером является компания Cloudflare, на карте связей интернет-сайтов (проекция «Alexa-Majestic»)

Если строить информационную карту связей между провайдерами и информационными ресурсами, то в ней будет более 100 000 узлов. Эти узлы образуют множество тематических, языковых и других видов кластеров, которые необходимо разметить. Однако, для анализа могут быть использованы построенные ранее карты (карта связей автономных систем, рис. 2 А и две проекции карты связей интернет-сайтов), которые образуют масштабируемую систему. Во-первых, как было показано на рис. 7, карты связей интернет-сайтов можно рассматривать как минимум в двух масштабах. Аналогичное утверждение имеет место и для карты связей автономных систем. Кроме того, совместное использование этих карт образует еще один уровень масштабирования. За счёт детализации объектов уровня автономных систем можно перейти к уровню информационных ресурсов (рис. 10).

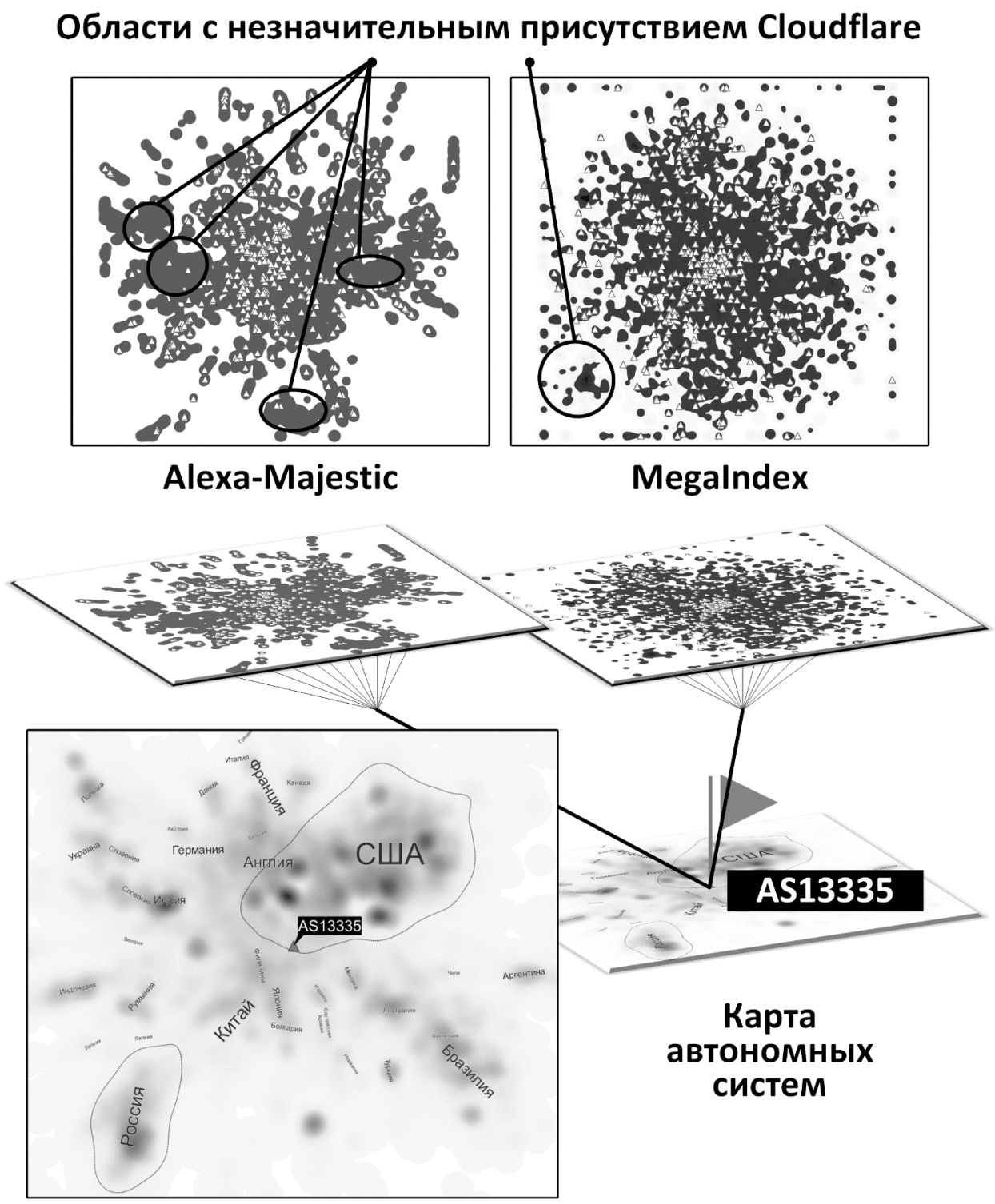

Рис. 10. Масштабируемая карта связей хотсинг-провайдеров и интернет-ресурсов, включающая две информационне карты объектов киберпространства, относящихся к различным информационным уровням (уровням «Автономные системы» и «Информационные ресурсы»)

В результате, путём детализации узла AS13335 (идентификатор автономной системы хостинг-провайдера Cloudflare), можно перейти как к проекции «Alexa-Majestic» (которая отражает языковые факторы), так и к проекции «MegaIndex» (демонстрирующей тематические зависимости).

Анализируя соответствующие карты, за несколько секунд можно установить то, что хостинг-провайдер предоставляет услуги самым крупным Интернет-сайтам в разных странах, а тематика таких ресурсов

охватывает практически все области. Данный вывод достаточно сложно сделать по табличному представлению или диаграмме, так как подобные формы предоставляют лишь факты или позволяют сравнить объекты, но не дают представления об общей картине (контексте, в котором находятся рассматриваемые объекты).

Таким образом, для построения генеральной карты защищаемого киберпространства на первом этапе необходимо сформировать набор информационных карт, изображающих как можно больше разнообразных слоёв киберпространства в различных проекциях. Это позволит выявить «скелет» киберпространства (граф связей уровней и слоёв киберпространства) и определить наилучшие способы его изображения (проекции для каждого слоя, которые наилучшим образом учитывают специфику состава и структуры его информационных объектов). Далее возможна детализация каждого из слоёв с учётом его положения в рамках генеральной карты. Для этого необходимо наличие многопользовательской информационно-картографической системы с достаточным объёмом вычислительных ресурсов, подобно существующим крупномасштабным геоинформационным системам [14, 20, 15], которые обеспечивают обработку геопространственных данных для всей планеты. В качестве аналогии процесса можно привести работы по формированию карты Вселенной, когда с помощью современных средств астрономии проводится изучение распределения вещества во Вселенной [22]. Такие работы позволяют выявить ранее неизвестные физические закономерности и законы. Того же можно ожидать и в случае построения карты киберпространства в отношении действующих в нём информационных законов.

В настоящий момент сформирована методическая и инструментальная основа для построения генеральной карты защищаемого киберпространства [23-27]. При этом задача построения такой карты требует совместных усилий всего сообщества, заинтересованного в проведении систематических исследований киберпространства в его различных аспектах. Для этих целей на платформе GitHub, создан аккаунт CyberMap [28], с помощью которого по мере развития проекта будут размещаться репозитории разработанных программных средств.

Список литературы

1. Браун Л.А. История географических карт // История ГИС от древности до XX века. М.: Центрполиграф. 2006. 479 с.

2. Владимиров В.Н. От исторического картографирования к исторической геоинформатике // Круг идей: алгоритмы и технологии исторической информатики: Труды IX конференции Ассоциации" История и компьютер". М. 2005. С. 22-40.

3. Grandjean M. Translating Networks: Assessing correspondence between network visualisation and analytics / M. Grandjean, M. Jacomy // Digital Humanities. 2019. 10 с.

4. Берлянт А.М. Картографический метод исследования // Изд-во Моск. ун-та. 1978. 257 с.

5. Робачевский А. Интернет изнутри: экосистема глобальной сети // М.: Альпина Паблишер. 2015. 223 с.

6. Язов Ю.К. Способ контроля эффективности ложной информационной системы, основанный на анализе широковещательных сетевых пакетов / Ю.К. Язов, А.Л. Сердечный, А.В. Бабурин // Информация и безопасность. 2013. Т. 16. Вып. 4. С. 506-509.

7. Hu Y. Traffic Classification of User Behaviors in Tor, I2P, ZeroNet, Freenet / Y. Hu, F. Zou, L. Li и др. // 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 2020. 418-424 p.

8. Jacomy M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software / M. Jacomy, T. Venturini, S Heymann b и др. // PloS one. 2014. Т. 9. №. 6. С. e98679.

9. Traag V.A. From Louvain to Leiden: guaranteeing well-connected communities / V.A. Traag, L. Waltman, N.J. Van Eck // Scientific reports. 2019. Т. 9. №. 1. С. 1-12.

10. Pochat V.L. Tranco: A research-oriented top sites ranking hardened against manipulation / V. L. Pochat, T. Van Goethem и

др. // arXiv preprint arXiv:1806.01156. 2018. С. 1-15.

11. Download Top 1 Million Sites. // URL: https://hackertarget.com/top-million-site-list-download/ (дата обращения: 09.09.2021).

12. Alexa Top 1 Million // URL: https://s3.amazonaws.com/alexa-static/top-1m.csv.zip (дата обращения: 09.09.2021).

13. Cisco Umbrella // URL: http://s3-us-west-1.amazonaws.com/umbrella-static/index.html (дата обращения: 09.09.2021).

14. Majestic Million // URL: https://majestic.com/reports/majestic-million (дата обращения: 09.09.2021).

15. Tranco // URL: https://tranco-list.eu/list/PGQJ/1000000 (дата обращения: 09.09.2021).

16. Common Crawl // URL: https://commoncrawl.org/the-data/get-started/ (дата обращения: 09.09.2021).

17. Megaindex // URL: https://ru.megaindex.com/ (дата обращения: 09.09.2021).

18. DomCorp // URL: https://www.domcop.com/top-10-million-websites (дата обращения: 09.09.2021).

19. Wikimapia // URL: https://wikimapia.org/ (дата обращения: 09.09.2021).

20. OpenStreetMap // URL: https://www.openstreetmap.org/ (дата обращения: 09.09.2021).

21. «Народная карта» Яндекса // URL: https://n.maps.yandex.ru/ (дата обращения: 09.09.2021).

22. Макаров Д.И. Построение карты близкой Вселенной : дис. Спец. астрофиз. обсерватория РАН, 2016.

23. Сердечный А.Л. Картографический подход к описанию киберпространства в контексте обеспечения безопасности информации и информационной безопасности / А.Л. Сердечный, А.Г. Остапенко // Информация и безопасность». Том 22. Ч.3.- Воронеж: ВГТУ, 2019. С. 387-398.

24. Сердечный А.Л. Концептуальные основы картографии защищаемого киберпространства. Часть 1. / Сердечный А.Л. // Информация и безопасность. 2021. Т. 24. Вып. 3. С. 373-386.

25. Сердечный А.Л. Концептуальные основы картографии защищаемого киберпространства. Часть 2. / Сердечный А.Л. // Информация и безопасность. 2021. Т. 24. Вып. 3. С. 387-400.

26. Сердечный А.Л. Информационно-картографические системы как инструментальная основа картографии защищаемого киберпространства / Системы управления и информационные технологии. 2021. № 4 (86). С. 41-46.

27. Сердечный А.Л. К вопросу о создании платформы картографирования рисков защищаемого киберпространства / А.Л. Сердечный, А.А. Гончаров, М.А. Булычев, А.В. Коноплин, О.С. Газизянов, Р.О. Дыкин, Д.С. Нестеров, Д.А. Нархов // Информация и безопасность. 2021. Т. 24. Вып. 4. С. 593-600.

28. Репозитории проекта CyberMap // URL: https://github.com/CyberMap (дата обращения: 09.09.2021).

Воронежский государственный технический университет
Voronezh State Technical University

Поступила в редакцию 12.09.2021

Информация об авторе
Сердечный Алексей Леонидович – канд. техн. наук, старший преподаватель, Воронежский государственный технический университет, e-mail: alex-voronezh@mail.ru

PROSPECTS FOR BUILDING THE GENERAL MAP OF PROTECTED CYBERSPACE MAPS

A.L. Serdechnyy

The article discusses the presentation of a general map of cyberspace at its physical, informational and social levels. With regard to the physical layer, the landscapes of operators, zones of interaction of IoT devices, and the location of cyber-physical systems are indicated. The nomenclature of information-level landscapes includes: data on logical routes, taking into account various network protocols; alternative protocols organizing the work of isolated virtual spaces; content maps of information resources, ontologies and data banks. At the social level, there is an abundance of objects with multiple interlayer intersections. The coordinates of objects in cyberspace are also considered through the identifiers of a level, a typical layer, a dataset, a layer projection algorithm, an object image within a layer. Cyberspace layers are proposed to be defined taking into account the scalability. Cyberspace layer projections allow viewing the same objects in different layouts on a plane or in three-dimensional space. Scaling the layers of cyberspace allows you to organize directed work with big data. The reader is offered numerous examples illustrating the above guidelines.

Keywords: general map, levels, identifiers, layers, objects, scaling, projections.

Submitted 12.09.2021

Information about the author
Alexey L. Serdechnyy – Cand. Sc. (Technical), Senior Lecturer, Voronezh State Technical University, e-mail: alex-voronezh@mail.ru