ИТ в помощь СБ

Как аналитику службы безопасности решить проблему анализа больших объёмов данных?

ДИРЕКТОР ПО БЕЗОПАСНОСТИ
№2 ФЕВРАЛЬ 2010

Андрей Майоров
руководитель отдела аналитических систем, компания РДТЕХ

Для решения спектра задач, стоящего перед аналитиками служб безопасности, проблема структурирования и систематизации больших объёмов данных является одной из главных. Очевидно, что в данном случае без средств автоматизации обойтись практически невозможно - ту часть работы, которую в состоянии быстро и без «человеческих» ошибок сделать программное обеспечение, необходимо доверить специализированным информационно-аналитическим системам (ИАС), созданным под потребности служб безопасности. Чтобы понять, какие именно требования предъявлять к таким системам, необходимо очертить круг типичных задач, которые в ходе своей работы решает работник службы безопасности:

- конкурентная (экономическая, бизнес) разведка;
- организация работы по правовой, организационной и инженерно-технической защите коммерческой тайны;
- проверка персонала;
- охрана предприятия.

Наиболее трудоёмкой аналитической работой, безусловно, является конкурентная разведка, т.е. сбор и обработка информации, потенциально влияющей на выработку управленческих решений.

К такого рода информации могут относиться как правовые акты, заявления крупных государственных чиновников, так и данные об экономическом положении конкурентов, структуре их бизнеса, аффиляционные связи между компаниями, данные о личных связях ключевых фигур.

Основными задачами, решаемыми в ходе конкурентной разведки, являются:

- изучение торгово-конъюнктурных ситуаций в пространстве деятельности учредителей, партнеров, клиентов и потенциально возможных конкурентов;
- ситуационный анализ текущего состояния финансово-торговой деятельности с точки зрения прогнозирования возможных последствий, могущих привести к неправомерным действиям со стороны конкурирующих организаций и предприятий;
- выявление платежеспособности юридических и физических лиц, их возможности своевременного выполнения платежных обязательств;
- установление антагонистических конкурентов, выявление их методов ведения конкурентной борьбы и способов достижения своих целей;
- определение возможных направлений и характера злоумышленных действий со стороны специальных служб промышленного шпионажа против предприятия, его партнеров и клиентов.

ТРЕБОВАНИЯ К ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЕ СЛУЖБЫ БЕЗОПАСНОСТИ

Решение этих задач определяет ряд специфических требований к программному обеспечению.

Аналитику СБ приходится работать довольно с широким набором типажей объектов и связей между ними. Более того, в ходе анализа данных может возникнуть необходимость определения новых, ранее отсутствовавших в системе типов для описания

объектов, привлекших внимание аналитика. Таким образом, семантическая модель аналитической системы СБ должна быть легко расширяема без участия разработчиков, если не самим аналитиком, то по крайней мере администратором системы.

Аналитическая деятельность СБ, как правило, не может быть ограничена некоторым отдельным внутренним источником данных. ИАС СБ должна быть способна использовать как собственные корпоративные информационные ресурсы, так и внешние. Внешние ресурсы - это открытые источники (публикации в Интернете и в традиционной прессе), данные информационных служб (например, LexisNexis или SPARK), предоставляемые по подписке, а также коммерчески доступные базы данных по различным тематикам. Здесь речь, как правило, идёт не о некоторой регламентированной процедуре загрузки некоторого заранее определенного набора данных, а скорее, о возможности формирования аналитиком запроса по объекту к внешним базам данных. Таким образом, программное обеспечение ИАС СБ должно предусматривать как возможность загрузки данных из разнородных источников, так и интеграцию разнородных баз данных в единую информационную систему.

В случае работы с открытыми публикациями речь, прежде всего, идет о возможности работы с неструктурированной информацией, а именно, автоматизации процесса анализа документов с выявлением в нём объектов интереса и связей между ними, а также сохранении этих объектов в базе данных в структурированном виде для последующего более детального анализа.

Множество внешних источников данных неизбежно порождает в ИАС дублирующие записи об одном и том же объекте. Наличие средств поиска похожих объектов является неотъемлемой частью программного обеспечения СБ.

Одной из основных особенностей анализа в СБ является акцентирование внимания на отношениях и зависимостях - связях между объектами. Такого рода свойство должно находить отражение как в логической модели базы данных, так и в графических средствах их отображения.

Широко применяемые структуры "мастер-деталь" в OLTP-базах данных или "звезда" - в OLAP-витринах не совсем удобны для моделирования данных, содержащих большое количество сущностей, и связанны друг с другом, как правило, отношениями "многие ко многим". При этом "многие ко многим" следует рассматривать не только в традиционном смысле, предполагаемом ER-моделью, но и в том смысле, что каждая сущность, как правило имеет отношения с множеством других сущностей. При этом очевидно, что один и тот же тип отношения может связывать не только две сущности, но и их произвольное количество. Например, сущность "юридическое лицо" может быть связано само с собой отношением "учредитель" ("многие ко многим"), и в то же время - таким же отношением с сущностью "Физическое лицо".

Для аналитиков, занимающихся следственной деятельностью, наиболее подходящей моделью представления данных является граф, т.е. набор объектов со связями между ними, поскольку значительная часть аналитической работы в этом случае заключается в выявлении связей между объектами.

Здесь на первый план выходят визуальные средства анализа и такие графические представления данных как:

- схемы связей;
- схемы последовательности событий;
- схемы транзакций.

Каждое из трех, обозначенных в рисунках основных представлений, наилучшим образом демонстрирует тот или иной аспект взаимоотношений исследуемых объектов. В первом случае - факт наличия прямых либо косвенных связей между объектами; во втором – временную последовательность общих событий, в которые были вовлечены объекты; в третьем – наличие и временное распределение транзакций между объектами, т.е. информационных, материальных или финансовых потоков.

Результат работы аналитика - это не просто поиск фактов, но и превращение их в информацию. Работник СБ должен иметь возможность оценить и зафиксировать качество и достоверность данных, добавить комментарии, возможно, какие-то связи к уже имеющимся. Например, предположим, что в базе данных мы нашли данные о продаже недвижимости

На схемах ниже - четыре лица, совершавшие продажу/покупку трех квартир. Сами данные ничего не говорят ни об отношениях между людьми, ни о характере эпизода. Только дополнительный анализ дат сделок, имен участников, их дат рождения позволяет аналитику установить, что указанные лица являются близкими родственниками, Чаплыгина и Баронкина - одно и тоже лицо, а сделка представляет собой размен 2-х комнатной квартиры на две однокомнатные, с разъездом старшего и младшего поколений.

Для аналитика чрезвычайно важно, чтобы программное обеспечение позволяло формулировать и фиксировать свои умозаключения, в том числе, и посредством визуальных средств на схеме. Расследование, как правило, длится довольно долго, и аналитик должен иметь возможность

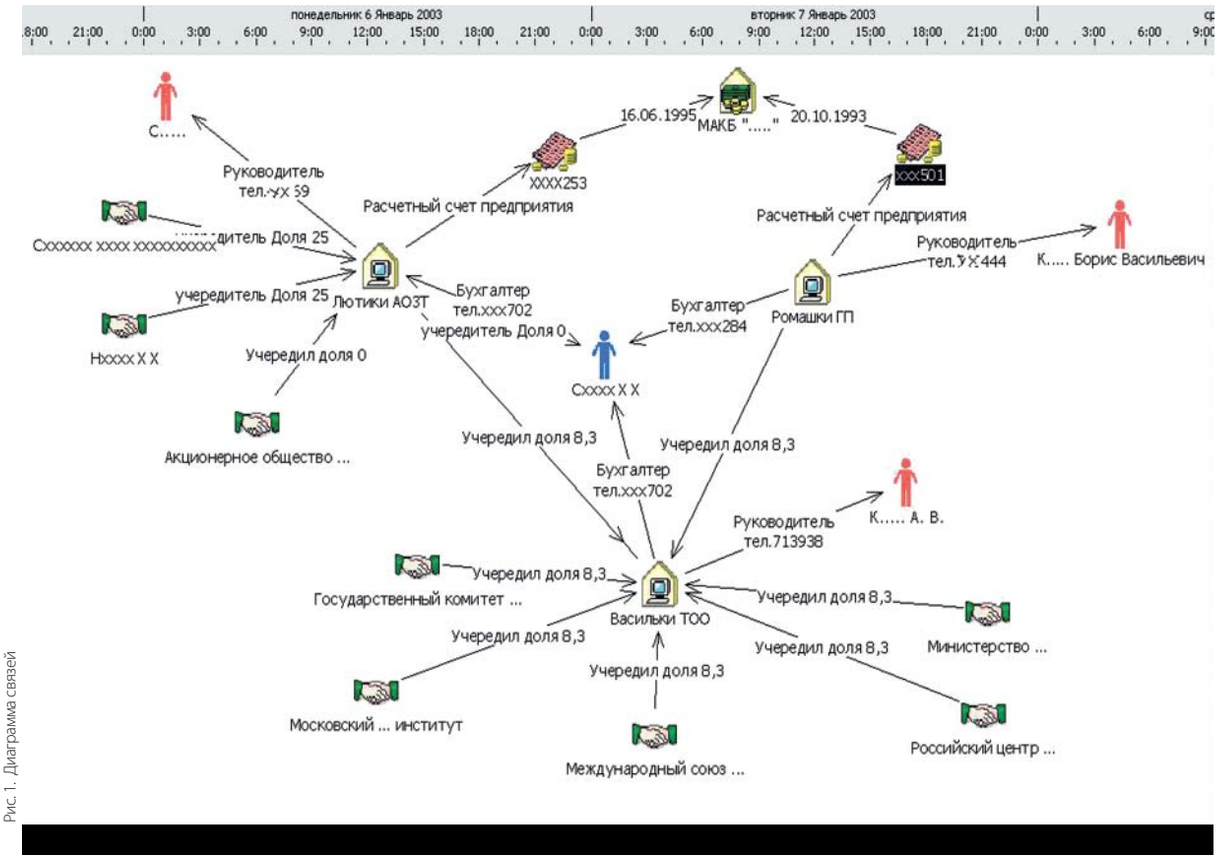


Рис. 1. Диаграмма связей

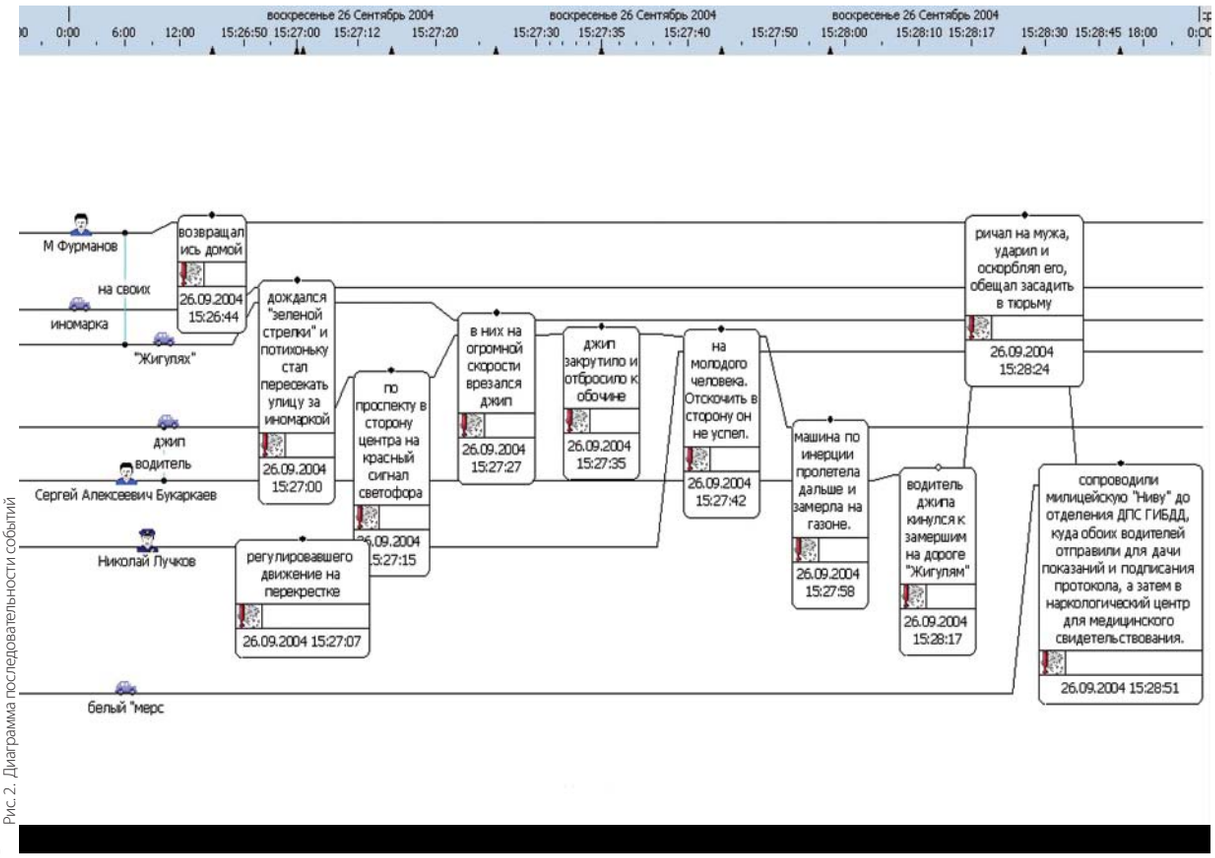


Рис. 2. Диаграмма последовательности событий

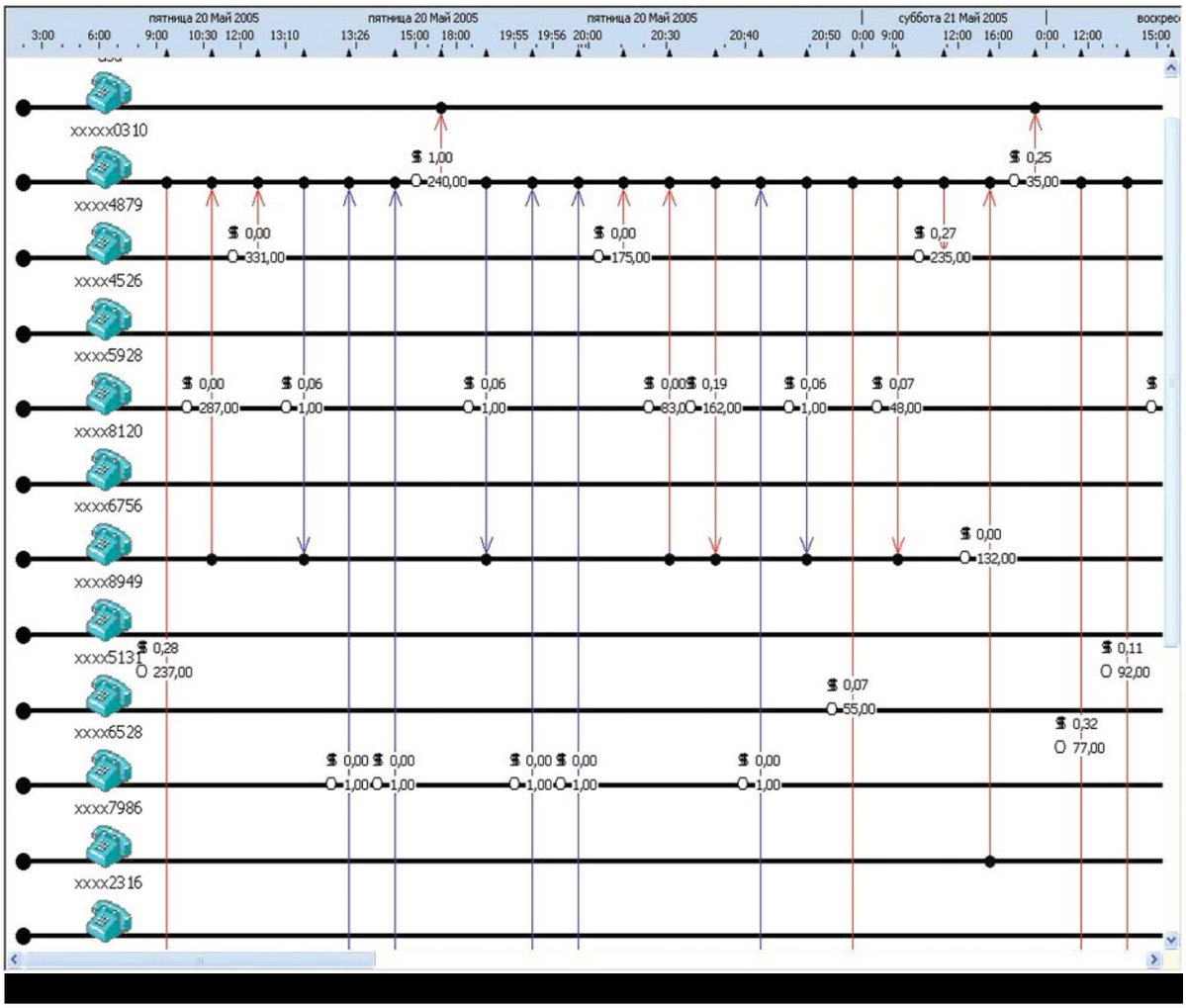


Рис. 3. Диаграмма транзакций

сохранять объекты расследования в «деле» с тем, чтобы не выполнять их поиск многократно в источниках данных.

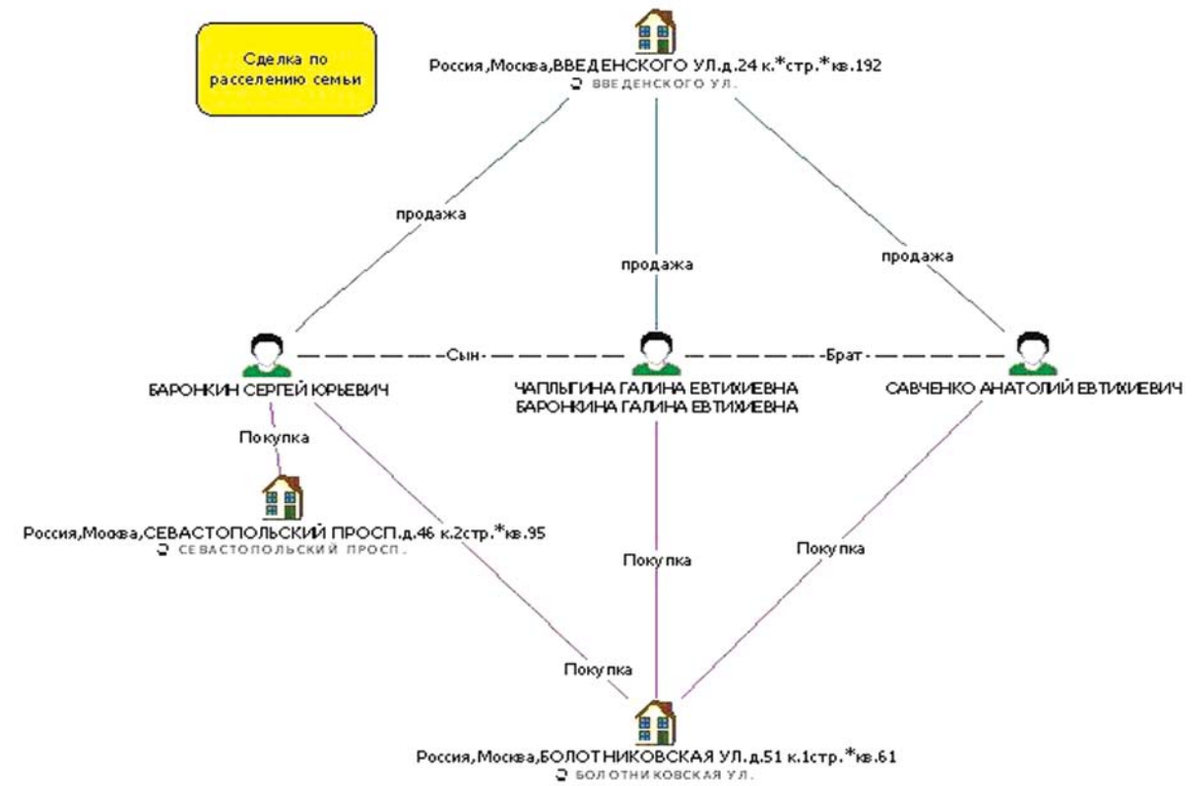
Результатам анализа является тот или иной документ, подготавливаемый аналитиком, а «дело» может быть источником данных для первоначального автоматически генерируемого шаблона документа. Не последнюю роль играют специализированные алгоритмы, используемые в анализе данных, такие как: кластеризация, поиск пути между объектами, поиск шаблонов поведения и т.п.

Таким образом, основополагающие требования к функционалу ИАС СБ можно сформулировать как:

- возможность работы с различными типажами объектов;
- акцент на выявлении связей и отношений объекта анализа с прочими объектами;
- работа с внешними источниками, как коммерческого характера, так и предоставляемых в качестве обмена прочими структурами;
- поиск дубликатов;
- работа с неструктурированной информацией;
- работа с «делом»;
- представление данных в ходе анализа, а также его результатов в виде диаграмм и схем;
- оценка качества и достоверности информации;
- формулирование умозаключений и выводов об объектах анализа;
- оформление результатов анализа в виде аналитических записок и отчетов;
- использование специализированных аналитических функций.

ЛИЦЕНЗИОННЫЕ ПРОГРАММНЫЕ ПРОДУКТЫ ДЛЯ СЛУЖБ БЕЗОПАСНОСТИ

На рынке программного обеспечения, ориентированного на задачи следственной деятельности,



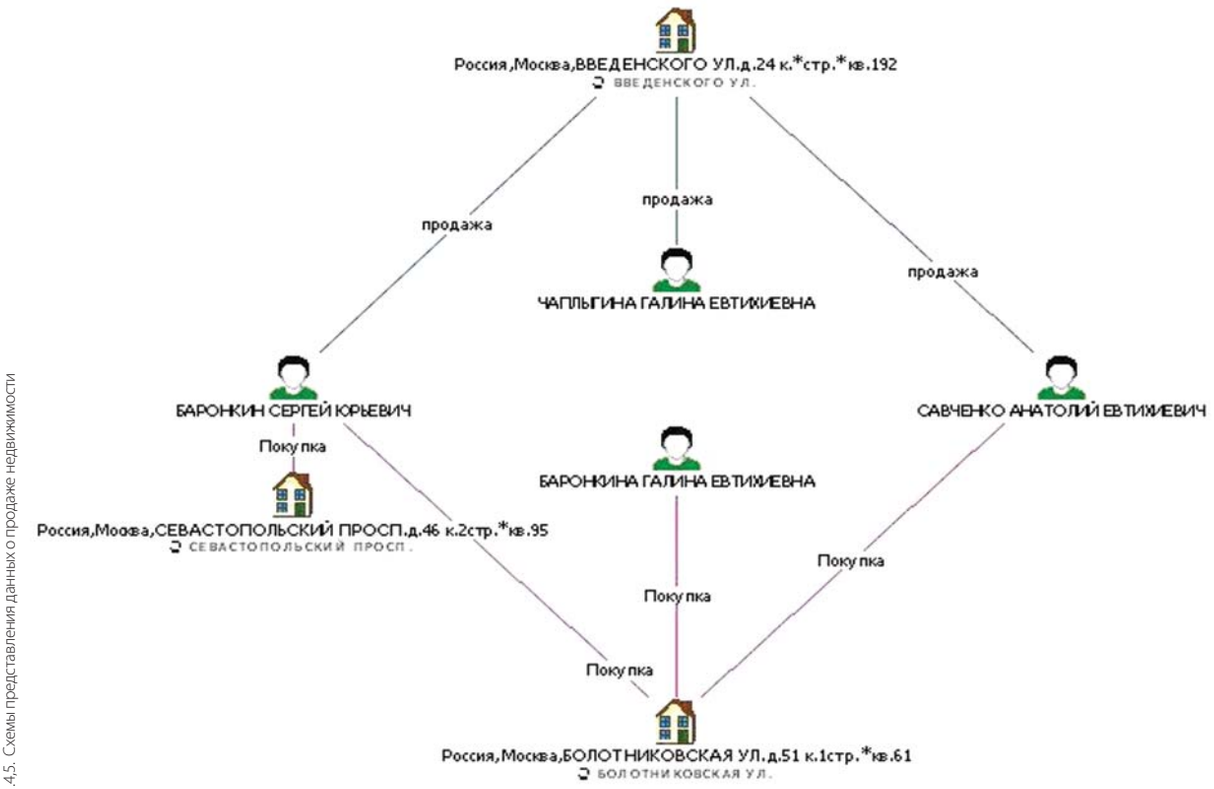


Рис.4,5. Схемы представления данных о продаже недвижимости

можно выделить две группы продуктов: коробочные продукты и разрабатываемые под заказ, либо требующие участия производителя в конфигурации продукта.

Первая группа представлена на российском рынке, главным образом, иностранными производителям - общепризнанными лидерами мирового рынка: i2 Limited, Visual Analytics Inc.

Российские компании в основном специализируются на кастомизированной разработке ИАС, используя как лицензионные продукты (в т.ч. упомянутых выше компаний), так и собственные разработки.

Компании i2 Ltd, Visual Analytics Inc. присутствуют на российском рынке информационных технологий уже много лет, и каждая имеет свой круг пользователей.

Флагманские продукты этих компаний, существенным образом отличаясь архитектурно, используют одну и ту же модель данных - "объект-связь-объект" и в значительной мере пересекаются функционально, поскольку так или иначе пытаются решить основные задачи:

- визуального анализа данных;
- хранения данных, появляющихся во время расследования;
- использования данных, хранимых во внешних базах данных;
- работы с неструктурированными данными.

Поскольку основной упор в программном обеспечении подобного рода делается на визуальное восприятие, то графические средства представления информации играют главную роль. Каждый из основных типов диаграмм, представленных на рисунках 1-3, имеет множество вариаций автоматического расположения объектов, подчеркивающих тот или иной аспект отношений объектов.

Безусловно, функциональность продуктов не ограничивается только графическим представлением данных. Современные программные средства для аналитиков СБ предоставляют пользователю широкий выбор команд, позволяющих более эффективно работать с данными. Среди наиболее интересных возможностей следует отметить:

- **Расширение связей**. Выделив на схеме объект, аналитик может одним щелчком мышки найти в базе данных не только прямые, но и косвенные связи данного объекта глубиной до пяти уровней.
- **Поиск пути между объектами: как на схеме, так и в базе данных**. Функция позволяет показать цепочку объектов и связей между ними. В случае поиска на диаграмме, у пользователя есть возможность задать самый короткий путь. Также, если связи несут информацию о дате и времени, - самый ранний путь. Более того, при необходимости можно задать учет направления связи, что актуально в случае анализа банковских транзакций.

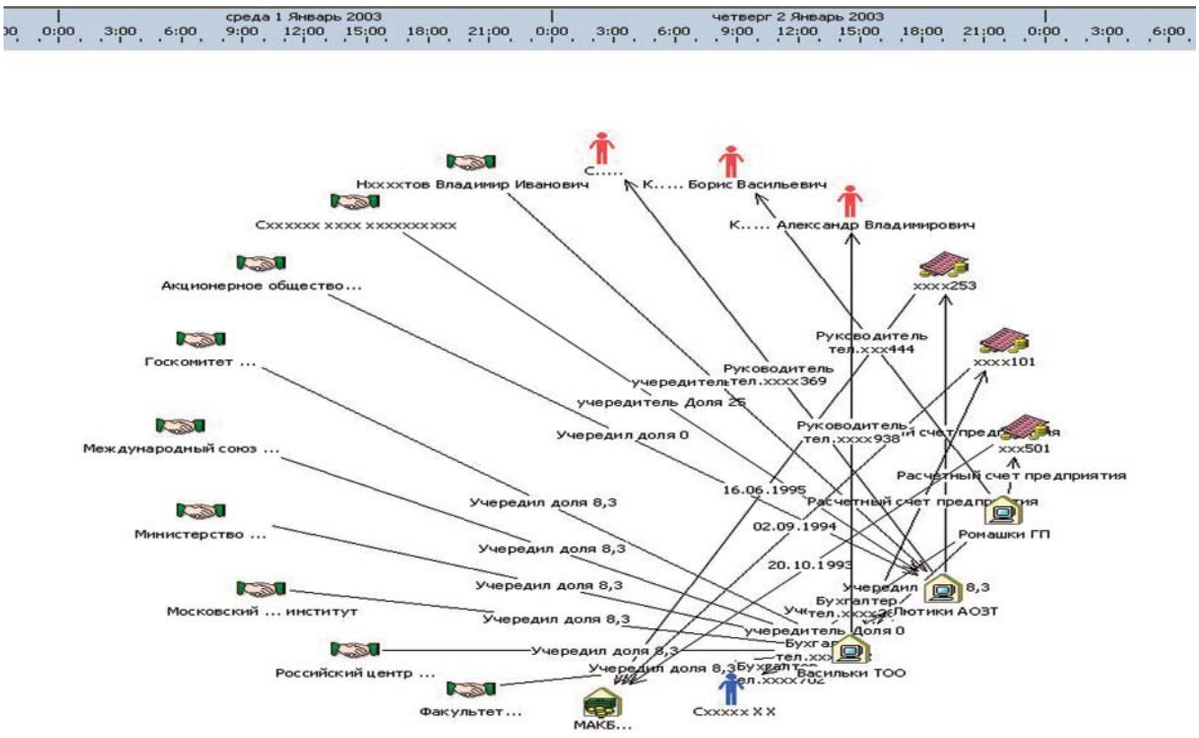


Рис. 6 Круговое расположение – объекты расположены по кругу с группировкой по типам и количеству связей

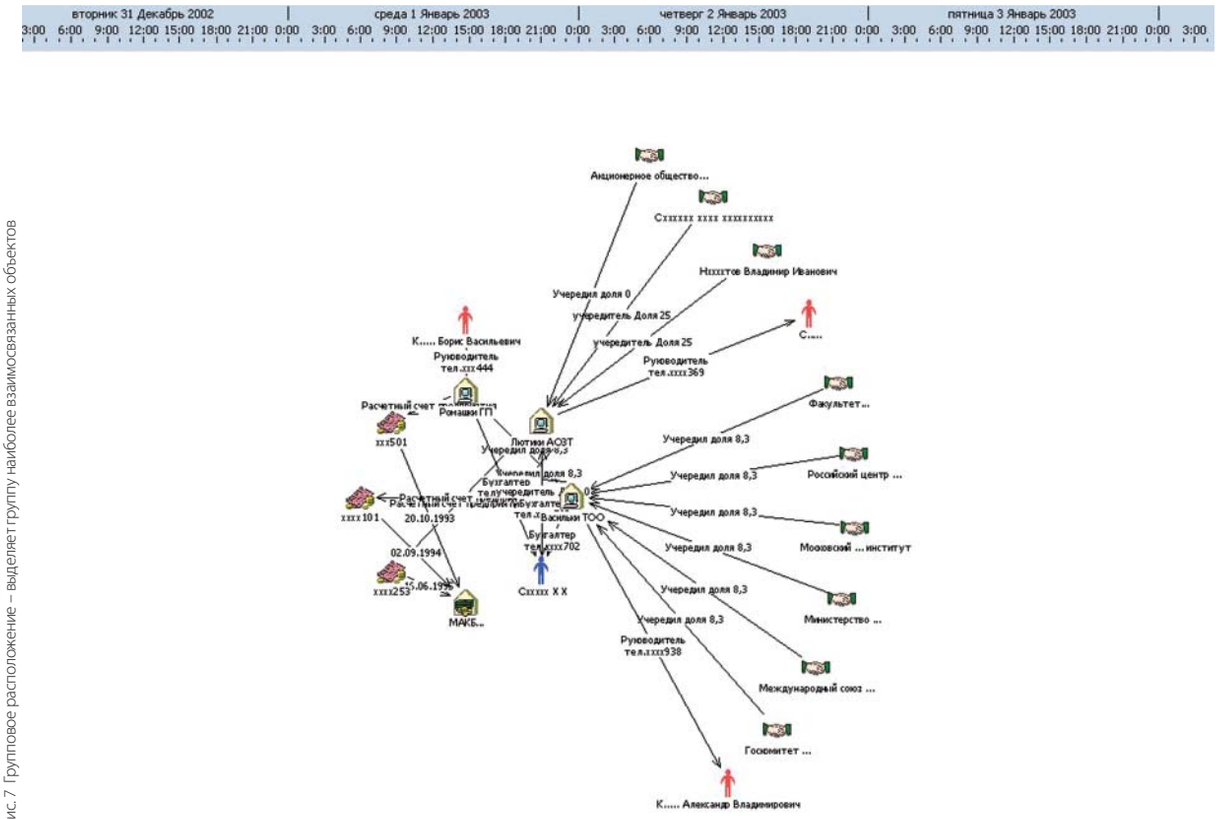


Рис. 7 Групповое расположение – выделяет группу наиболее взаимосвязанных объектов

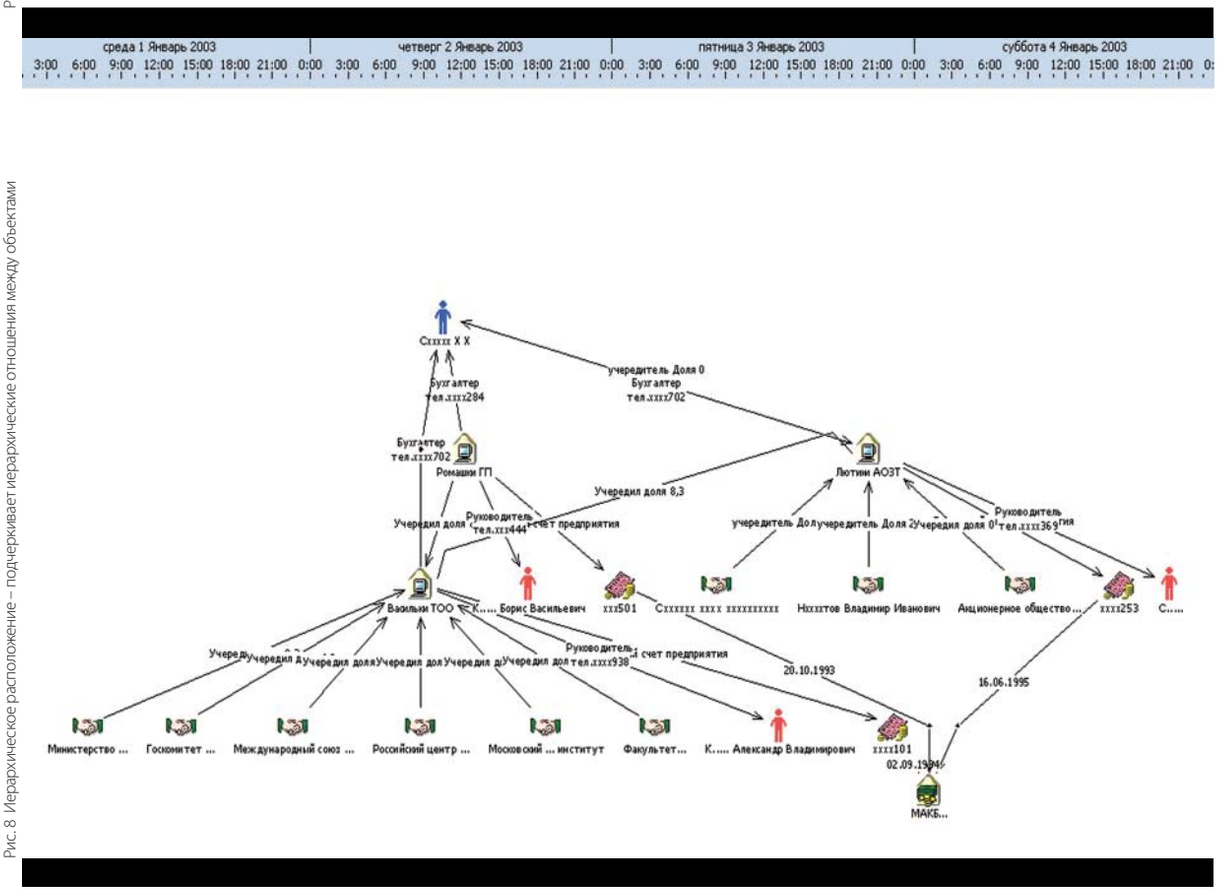


Рис. 8 Иерархическое расположение – подчеркивает иерархические отношения между объектами

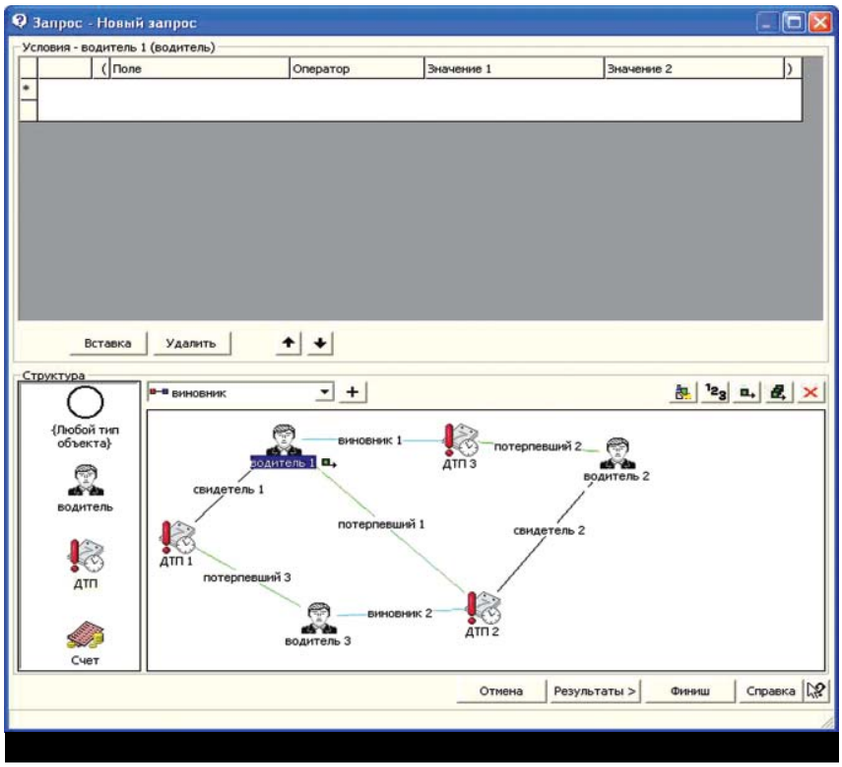


Рис. 9 Пример шаблона события

- **Кластеризация на основе взаимных связей группы объектов.** Возможность выделять на схеме группу наиболее тесно связанных между собой объектов.
- **Поиск похожих объектов**. Группировка объектов со схожими именами, автоматическая идентификация дубликатов объектов на схеме до определенной степени, что позволяет решить проблему консолидации данных об объекте, хранимых в разных базах.
- **Широкие возможности поиска объектов: и на диаграмме, и в базе данных.** Помимо простейшего поиска объекта заданного типа по значениям атрибутов, имеется возможность генерации графического запроса. Данная функция особенно интересна, поскольку позволяет аналитику создавать шаблоны событий. Так, например, аналитику страховой компании это может быть актуально для поиска группы водителей, одновременно вовлеченных в серию ДТП, но в разных ролях, как показано на рисунке 8, где группа мошенников в различных ДТП поочередно выступает, то в роли потерпевшего, то в роли виновного, то в роли свидетеля.

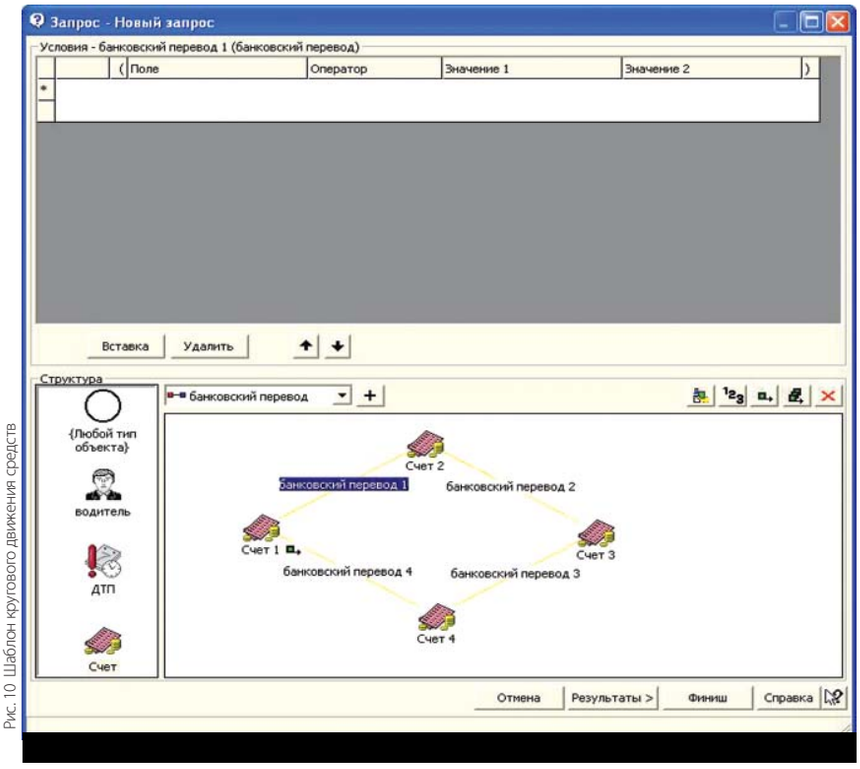


Рис. 10 Шаблон кругового движения средств

Аналитик, занимающийся выявлением фактов отмывания денег, может использовать запрос, графически изображенный на рисунке 10 для поиска случаев кругового движения средств в группе счетов.

Таким образом, основные достоинства современного программного обеспечения для СБ можно охарактеризовать так: это широкие возможности по поиску и структурированию больших объёмов данных, поступающих из разных источников, способность визуализации любой информации и ориентированность на все типы задач, стоящих перед аналитиками служб безопасности.