УДК 004.7.052:004.414.2 **Коробко В.В., Скоропаденко А.А.**

МЕТОДЫ ПОСТРОЕНИЯ КОМБИНИРОВАННЫХ СИСТЕМ ОБНАРУЖЕНИЯ АТАК НА ЦИФРОВЫЕ СЕТИ

Введение

В настоящее время разработкам в области обеспечения информационной безопасности цифровых инфокоммуникационных сетей уделяется повышенное внимание. Действующие специализированные департаменты в корпоративном секторе имеют одни из наибольших бюджетов по сравнению с другими структурными подразделениями. Наблюдаются закономерные тенденции оптимизации времени реагирования и комплексов контрмер к соответствующим угрозам, а также по повышению эффективности управления и контроля комплексных систем безопасности. При этом увеличение количества разновидностей потенциальных угроз, а также интенсивность их возникновения приводят к необходимости улучшения существующих и разработок новых методов и планов реагирования [1,8]. Разрабатываются общие и специальные рекомендации, методики ранжирования типов угроз информационной безопасности.

Для обеспечения оперативного контроля и прогнозирования состояния информационных ресурсов необходимо создание новых технологий обнаружения признаков удаленной атаки с использованием активных и пассивных методов и датчиков слежения, выходящих за рамки рутинного аудита операционных систем и журналов сетевых протоколов.

Состояние проблемы защиты сетей от несанкционированных вторжений

Для создания эффективной методики защиты необходима разработка системы профилирования массива данных, несущих информацию об угрозе. Существующие системы обнаружения сетевых атак [2] достаточно эффективно выполняют детализированные расследования и поиск злоумышленников на основании полученных во время атаки данных, однако, они не в состоянии прогнозировать цели, намерения, дальнейшие действия злоумышленников. В частности наиболее распространенные системы обнаружения вторжений, которые используются для обеспечения безопасности систем, обрабатывающих данные открытого характера, функционируют посредством сравнения элементов потока IP-данных с имеющейся базой сигнатур. На основании сравнительного анализа система принимает решение блокировать тот или иной пакет, либо пропускать для взаимодействия непосредственно с операционной системой. Сигнатурные системы обнаружения вторжений обладают высокой производительностью, эффективностью обнаружения при сравнительно невысоких требованиях к аппаратному обеспечению [3]. Однако они имеют ряд недостатков:

- невозможность ввода новых сигнатур;
- отсутствие системы блокирования неизвестных сигнатур;
- отсутствие возможностей прогнозирования действий злоумышленника;
- отсутствие подсистемы мониторинга аппаратных ресурсов.

Учитывая большое количество данных, разнящихся как по качественным, так и по количественным характеристикам, целесообразно рассматривать некоторые последовательности событий в виде регулярных конфигураций, что, в свою очередь открывает следующие возможности:

- обнаружение сходства элементов, образующих конфигурации;
- обнаружение сходства конфигураций;
- обнаружение условий, при которых непроизводные элементы могут взаимодействовать между собой;
- изучение внутренней топологии регулярных структур.

Сетевые вторжения, локализуемые при помощи анализа соответствующих сигнатур в потоке данных, представляют собой массив несвязанных между собой конфигураций. Однако они включают элементы, которые можно объединить по определенным признакам. В рамках категорий сетевых вторжений (сбор информации, попытка несанкционированного доступа, отказ в обслуживании, подозрительная активность, системные атаки), существуют подвиды атак, которые могут быть реализованы с большей или меньшей вероятностью в зависимости от операционной системы, коммутационного оборудования, системы обнаружения вторжений, типа искомых данных, организационной инфраструктуры и т.д. Однако такое группирование не позволят делать выводы, необходимые для всестороннего изучения и принятия адекватных контрмер.

Постановка задачи

Существующие сигнатурные системы обнаружения вторжений [4] не могут профилировать перехваченный поток данных распределенных атак для их классификации и отработки сигнала о вторжении.

Второй тип систем обнаружения атак – системы, реагирующие на аномалии в сети как на протокольном, так и на программном уровнях, – не могут адекватно реагировать на разнообразные атаки. Кроме того, из-за сравнительно высоких вероятностей ошибок первого и второго рода их работа в режиме реального времени на объектах инфраструктуры, критичной к атакам, оказывается малоэффективной.

Учитывая вышеизложенное, можно сделать вывод, что для вынесения предположения о цели и последствиях атаки необходима разработка системы обнаружения вторжений, профилирующей комбинации сигнатур и сетевых аномалий на основании задаваемых признаков. Инструментом для эффективного группирования и обработки может стать методика упорядочивания данных, которые не могут быть классифицированы по непосредственным признакам. Наиболее перспективным направлением решения данной задачи является теория распознавания образов [5-7].

Математическая модель

Система распознавания является сугубо специализированной системой, предназначенной для работы в составе системы защиты сети. Разработка системы распознавания связана с решением определенной последовательности задач.

Первой задачей является детальное изучение объектов, которые подлежат распознаванию. Степень детальности этого изучения зависит от нужной эффективности работы системы распознавания (вероятностей принятия правильных и ложных решений). Результатами изучения объектов распознавания должны быть обоснованный выбор принципа классификации и определения количества объектов. Второй задачей является составление некоторого словаря признаков, которые используются для априорного описания классов объектов, которые подлежат распознаванию. Признаки объектов могут быть логическими и описываться детерминированными качественными выражениями или количественными величинами, которые попадают в конкретный интервал значений. Если признаки являются чисто случайными и распределенные по некоторому (в общем случае – неизвестному) числу классов или по всем классам объектов с определенным законом распределения, нужно использовать статистические методы. Признаки объектов, которые подлежат распознаванию, следует трактовать как вероятностные также тогда, когда результаты измерений их числовых значений получены с такими ошибками, что за этими результатами невозможно отнести объект к тому или другому классу с приемлемым качеством. Поэтому логично считать, что некоторые признаки будут отнесены к детерминированным или к стохастическим в зависимости от априорных данных и качества измерительным приборам. Если априорная плотность вероятности того или другого признака

имеет настолько малую дисперсию, что распределение можно считать дельтообразным (с допустимой для данного инженерного применения точностью), такой признак следует отнести к логическим (детерминированных). Такое предположение можно сделать также тогда, когда устройство для измерения количественных параметров признака имеет настолько высокую точность, что вероятность ложного отнесения результата измерений близка к нулю.

На практике чаще всего имеет место ситуация, когда признаки объектов разных классов частично перекрываются. В этом случае даже для признаков логического (детерминированного) типа невозможно достоверно, с вероятностью, равной единице, различить классы объектов, имеющих такие признаки. Соответственно, нужно использовать статистический подход.

Поскольку нас интересуют классификационные признаки объектов, порождаемых альтернативными распределениями вероятностей достаточно решить лишь вопрос о том, какая из плотностей вероятностей больше в данной точке, которая соответствует классу признаков. Поэтому наиболее целесообразно применять следующие решающие правила:

- основанное на методе *k* ближайших соседей;
- основанное на методе гистограмм;
- основанное на разложении по базисным функциям.

Некоторые из упомянутых методов применялись при решении задачи оптимального распределения ресурсов защиты телекоммуникационной сети от разных типов несанкционированных вторжений, включая и физические [10].

Статистические системы распознавания как системы проверки сложных гипотез против сложных альтернатив состоят из модульных логических элементов, называемых образующими. Любая образующая обладает неотделимыми от нее связями, которые могут быть ориентированными (вход $\rightarrow$ выход) или неориентированными. Образующая, имеющая только входные и/или выходные связи, называется ориентированной, а образующая со связями произвольного направления – неориентированной. Образующая представляется набором символов, который называется вектором признаков образующей.

Системы распознавания строятся из связки двух или большего числа образующих. Каждая связка системы распознавания, в зависимости от данных, присвоенных паре ее связей, может находиться в одном из двух состояний - истинном (замкнутом) или ложном (разомкнутом). Путем замыкания и размыкания связок в системе распознавания, составленных из ориентированных образующих, моделируют соединения и разъединения выходов и входов модулей, из которых состоят реальные системы.

Логические и физические модули с входами и выходами можно представить векторами ориентированных образующих:

$$a(gi) = a(i, \gamma_{il}, \beta_{im}^{il}, \beta_{ir}^{out}) . \qquad (1)$$

Компоненты $\gamma_{il}, \beta_{im}^{il}, \beta_{ir}^{out}$ параметрической образующей g_i делятся на две группы. Компоненты первой группы, представленные символами с нижними индексами, называются атрибутами образующей. Если l=1, то образующая gi имеет только один атрибут γ_{il}. Компоненты второй группы, представленные в векторе (1) символами β, называются в общей теории распознавания образов показателями связей образующей. В теории распознавания образов [8] показатели связей β и атрибуты γ вектора (1) трактуются как переменные, имеющие соответствующие области значений.

Параметры l, m, r, фигурирующие в нижних индексах переменных и образующей gi, могут принимать различные числовые значения:

l=1,2,...; m=0,1,2,...; r=0,1,2,.... В результате изменения значений параметров m, r из вектора (1) получаются векторы компонент образующих с разными числами входных и выходных связей.

Образующими, которые определяются вектором (1), представляют структуры реальных модулей в обобщенной форме. Поставим в соответствие переменным $\gamma_{il}, \beta_{im}^{il}, \beta_{ir}^{out}$ множества

$$D_{il}, D_{im}^{in}, D_{ir}^{out}, \tag{2}$$

называемые доменами. В доменах помещаются данные, присваиваемые переменным γ и β векторов компонент образующих.

Для получения образующих с разными числами входных и выходных связей параметры m, r в векторе (1) заменяются конкретными числами. Одновременно эти параметры заменяются в доменах (2) такими же числами. Вектор (1) и его домены (2) представляют собой образы структур и содержаний многих образующих, имеющих различные числа входных и выходных связей.

Особыми являются случаи, когда параметры m и r принимают значения, равные нулю. В случае, когда m=0, переменная β_{im1}^{in} и домен D_{im}^{in} исключаются из соотношений (1-2). В случае, когда r=0, из соотношений (1-2) исключаются переменная β_{ir}^{out} и домен D_{ir}^{out}. Таким способом строятся модели образующих, которые не имеют входных и выходных связей.

В дискретной теории распознавания образов применяются три вида образующих - абстрактные, конкретные и ассоциированные.

Абстрактная образующая определяется следующим образом. Во все домены образующей помещается неопределенное значение данных, обозначаемое символом λ_0. Образующая называется абстрактной, если во всех ее доменах содержится только символ λ_0 и ни в одном из них нет конкретных данных, характеризующих реальные модульные объекты. Следовательно, абстрактная образующая не определена на какой-либо конкретной информационной среде.

Образующая, в доменах которой, помимо символа λ_0, помещены данные об одном или нескольких реальных модулях, называется конкретной. Абстрактная образующая превращается в конкретную после размещения в ее доменах данных о реальных модулях. Конкретные образующие занимают промежуточное положение между абстрактными и ассоциированными образующими. Домены конкретных образующих определяются в общем случае как конечные или счетные множества значений переменных γ и β. В этом они аналогичны доменам атрибутов реляционных отношений, которые определяются в теории реляционных баз данных как конечные или счетные множества значений атрибутов.

Если переменным γ и β конкретной образующей присваиваются взятые из доменов данные о реальном модуле, то образующая становится ассоциированной с данными и служит классификационной моделью этого модуля.

При анализе событий, которые могут классифицироваться как сетевые вторжения, образующими, согласно теории распознавания образов, являются дейтаграммы сетевого трафика, сигнатуры как комбинации направленных дейтаграмм – регулярными конфигурациями, комбинации сигнатур и сетевых аномалий – изображениями, а профиль атаки в терминах теории распознавания образов – образом.

Известно, что система сетевой безопасности имеет иерархическую структуру, представленную на различных уровнях (см. рис.1). Каждый уровень такой системы имеет свои собственные элементарные единицы, которые могут сочетаться определенным образом. Особую сложность представляет собой согласование переходов с уровня на уровень.

Для перехода от одного уровня к другому недостаточно просто определить правила выделения образующих каждого уровня. Необходимо, кроме того, определить правила соединения образующих в регулярные конфигурации [5]. Ясно, что если имеется конфигурация

$$K_i = \left\{O_1^i, O_2^i, ..., O_n^i\right\}, \tag{3}$$

которая получена, например, при использовании правила Li при объединении образующих Oij в регулярную конфигурацию, то она может быть описана в виде

$$K_i = \left\langle N_{k_i}, n, L_i \right\rangle, \tag{4}$$

где N_{k_i} – наименование конфигурации;

n – количество образующих в конфигурации;

L_i - правила сочетания образующих в конфигурации.

Выражение (4) является компактной формой записи (3). Таким образом, при описании уровней необходимо определение образующих и правил [7].

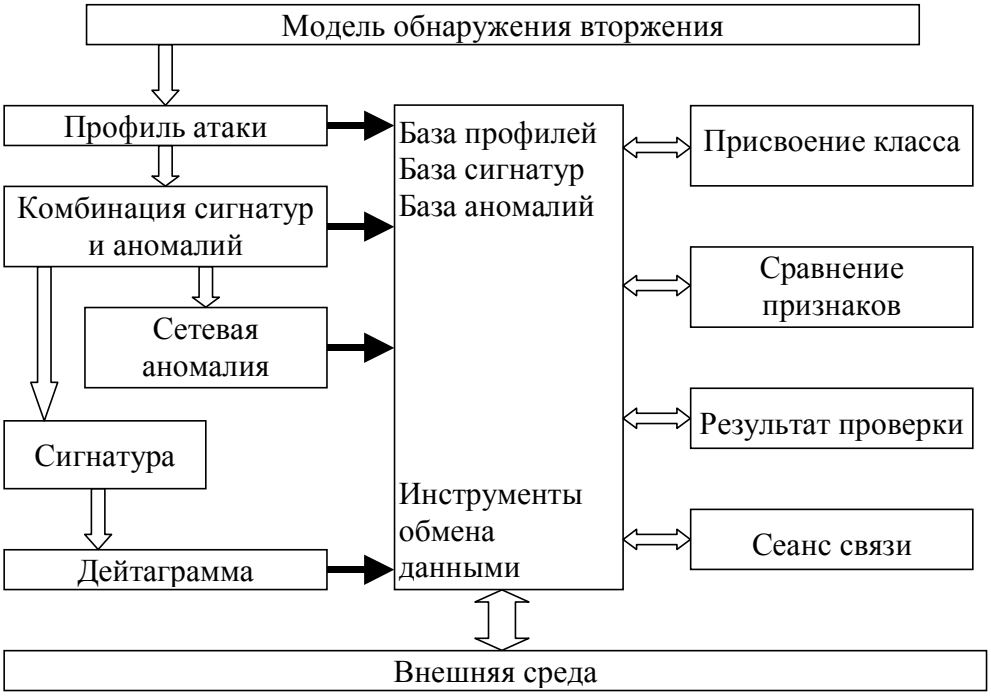

Рис. 1. Иерархическая структура модели обнаружения вторжения

Для выделения элементарных объектов физического уровня чаще всего используют предварительное выделение квазиоднородных участков сетевого потока. Будем считать, что множество квазиоднородных участков S={si}. Для выделения из потока элементарных объектов можно использовать пространство состояний критериев сложной системы. Для каждого объекта можно определить свое тело кластера Ki. Используя подходящую (квадратичную или по модулю расстояния) меру близости [7,9] в данном пространстве, можно определить образ, которому соответствует искомый квазиоднородный участок сетевого потока. Сравнения на данном уровне выполняются по определенным алгоритмам. Ясно, что алгоритмом сравнения, результатом работы которого являются образующие определенного уровня, в существенной мере определяются возможные операции. С другой стороны, алгоритмы могут быть построены на основе множества операций [9].

Определение отношений в результатах алгоритмов сравнения на заданном пространстве – достаточно сложная задача. С одной стороны, необходимо выразить численные параметры потока данных (период длительности, количество в единицу времени, интенсивность, временные характеристики, географические характеристики). С другой стороны, для кластеризации необходимо задать правила отнесения к какому-либо классу квазиоднородных участков.

Определим отношение эквивалентности R=~ двух квазиоднородных участков. Два квазиоднородных участка s1 и s2 будут эквивалентными, если будут принадлежать одному кластеру/классу в пространстве состояний, т.е. $s_1 \in K$ и $s_2 \in K$. Следовательно, все квазиоднородные участки, расположенные в одном кластере, принадлежат одному и тому же классу эквивалентности [K~]. Так классифицируются квазиоднородные участки. Отношение эквивалентности является рефлексивным, транзитивным, симметричным по определению.

Используя отношение эквивалентности, можно сделать отображение на множество наименований $M_N=\{m_i\}$ классов эквивалентности, m: $S \to M_N$, mi – наименование кластера класса эквивалентности [K~i]. У образующих наименований кластеров можно определить характеристики пространства состояний.

Определим следующую функцию $p(m_1, m_2)$: $M_N^2 \to [0 .. 1]$, где p – вероятность (частость) сочетаемости объектов m_1 и m_2 между собой. Значение 0 соответствует отсутствию сочетаемости. Значение 1 – объекты сочетаются между собой всегда. Таким образом, при определении регулярности той или иной конфигурации может быть использовано два подхода:

- для оценки регулярности использовать комбинации образующих (3), что наблюдается на верхних уровнях иерархии систем обеспечения сетевой безопасности.
- использовать описания правил регулярности конфигураций образующих (4), присутствующей в большей степени на нижних уровнях иерархии.

Оценка эффективности логической системы распознавания

При построении логических систем распознавания (к которым относится и рассматриваемая система), часто приходится сталкиваться с такой ситуацией. Значения истинности множеств наименований $M_N=\{m_i\}$ классов эквивалентности, которыми выражаются признаки классифицируемых объектов [K~], связанных соотношениями вида $s_i \in K, i = 1,2,\ldots,n$, устанавливаются по результатам предварительно собранных экспериментальных данных или непосредственно в процессе функционирования системы. Эти значения, как правило, устанавливаются не достоверно («да» – «нет»), а с известной неопределенностью, нечеткостью. Причинами этого могут быть следующие факторы.

1. Нечеткие высказывания типа «истинное значение δ лежит в пределах интервала $\Delta \ldots$», причем результат измерения значения δ есть случайная величина.

2. Непреднамеренные помехи и/или противодействие противника, вследствие чего создаются предпосылки для ошибочных заключений.

В рассматриваемой задаче наиболее вероятной причиной нечеткости является вторая.

Предположим, что выбранный способ описания классифицируемых объектов позволяет различать между собой все N классов как при представлении их с помощью логических функций $\varphi_N(s_i) = \begin{cases} 0 \\ 1 \end{cases}$, так и через предварительно измеренные или наблюдаемые в ходе опыта признаки $\tilde{s}_i = s_i + \delta s_i$, $i = \overline{1,N}$, δs_i – ошибки измерения (наблюдения). Для количественного описания искажений информации о классифицируемых объектах и вытекающих из этого ошибочных решениях, рассмотрим следующие вероятности:

$$\eta_i(1|1) = P\left(\tilde{s}_i = 1 \,|\, s_i = 1\right) \tag{5}$$

- вероятность того, что событие s_i действительно имело место, и будет принято решение $\tilde{s}_i = 1$;

$$\eta_i(\times|1)=P\left(\tilde{s}_i=\times|s_i=1\right) \tag{6}$$

- вероятность того, что событие s_i имело место, а значение истинности элемента $\tilde{s}_i$ не будет установлено;

$$\eta_i(0|1)=P\left(\tilde{s}_i=0|s_i=1\right) \tag{7}$$

- вероятность того, что событие s_i имело место, а будет принято решение $\tilde{s}_i=0$.

Аналогичные по смыслу вероятности $\eta(0|0),\eta(\times|0),\eta(1|0)$ можно записать и для случая, когда событие s_i не имело места.

Очевидно, события, описываемые вероятностями (7-9), составляют полную группу. Поэтому $\eta(1|1)+\eta(\times|1)+\eta(0|1)=1$.

Обозначим через $C_{ij}, i,j=1,2,\ldots\ldots,N$ выигрыш (при $i=j$ - отнесение $i-$ го объекта к $i-$ му классу) или штраф (при $i\neq j$ - отнесение $i-$ го объекта к $j-$ му классу), а через $C_{N\times,j}-$ штраф, который накладывается, если для объекта из $j-$ го класса не удается получить определенного решения.

Пусть $p_{ij}-$ условная вероятность принятия того или иного решения (5-7) о принадлежности обнаруженного объекта к $i-$ му классу при условии, что в действительности объект принадлежит к $j-$ му классу. Тогда условный средний выигрыш R_j от принятия решения при условии, что классифицируемый объект принадлежит к $j-$ му классу, определяется следующим выражением:

$$R_j=\sum_{i=1}^{N}C_{ij}\,p_{ij}+C_{N\times,j}\left(1-\sum_{i=1}^{N}p_{ij}\right)=C_{N\times,j}+\sum_{i=1}^{N}p_{ij}\left(C_{ij}-C_{N\times,j}\right). \tag{8}$$

Положим априорную вероятность появления объекта, принадлежащего к $j-$ му классу, равной F_j. Тогда безусловный средний выигрыш R от принятия решения в системе распознавания определяется выражением

$$R_j=\sum_{i=1}^{N}F_j\,R_j=\sum_{i=1}^{N}F_j\,C_{N\times,j}+\sum_{j=1}^{N}\sum_{i=1}^{N}F_j\,p_{ij}\left(C_{ij}-C_{N\times,j}\right). \tag{9}$$

Таким образом, разрабатывая различные алгоритмы определения истинности признаков s_i, и, соответственно, способы задания вероятностей (5-7), можно сравнивать эффективность этих алгоритмов по показателю (9). Выигрыши (или штрафы) C_{ij}, $C_{N\times,j}$, и априорные вероятности F_j первоначально задаются одинаковыми (из условия нормировки), а затем корректируются по мере накопления апостериорных данных в процессе функционирования системы распознавания.

Выводы

Аппарат теории распознавания образов целесообразно применять для построения систем профилирования сетевых вторжений.

Для выбора и обоснования адекватных моделей сетевого вторжения необходимо модифицировать известные методы обработки данных с адаптацией к конкретной задаче. Такой подход позволит создать основу для разработки системы профилирования данных, несущих информацию об атаке или угрозе вторжения.

Список литературы

1. Cannady, J. and J. Harrell. “A Comparative Analysis of Current Intrusion Detection Technologies.” 4 th Technology for Information Security Conference (TISC’96), May 1996.
2. Anita K. Jones and Robert S. Sielken. "Computer System Intrusion Detection" Survey Department of Computer Science, University of Virginia, September, 2000
3. Teresa F. Lunt. "A Survey of Intrusion Detection Techniques". Computers &Security. 12(4), June 1993.
4. Lunt, T.F. “Detecting Intruders in Computer Systems.” 1993 Conference on Auditing and Computer Technology, 1993.
5. Гренандер У. Лекции по теории распознавания образов. Синтез образов. / под. Ред. Ю.Журавлева; пер. с англ. - М.: Мир, 1979. – 383 с.
6. Вопросы статистической теории распознавания. / Барабаш Ю.Л., Варский Б.В., Зиновьев В.Т., Кириченко В.С., Сапегин В.Ф. – М.: Сов. радио, 1967.
7. Дуда Р., Харт П. Распознавание образов и анализ сцен. – М.: Мир, 1976. – 511 с.
8. Федорик С.И. Обнаружение сетевых вторжений методами теории распознавания образов (теории паттернов) // ПРОБЛЕМИ ІНФОРМАТИЗАЦІЇ ТА УПРАВЛІННЯ: Збірник наукових праць: Випуск 10. – К.: НАУ, 2004. – с. 113 – 119.
9. Маркус С. Теоретико-множественные модели языков. - М.:Наука, 1970, - 332с.
10. В.Ю. Скуйбида, В.В. Коробко, А.А. Скоропаденко, Я.В. Милокум. Оптимальное распределение ресурсов защиты сети передачи данных.

Поступила 8.12.09

УДК 681.3.06 **Емельянов С.Л.**

СУЩНОСТЬ И МЕТОДЫ КОМПЬЮТЕРНОЙ РАЗВЕДКИ

Введение

Возрастающее по экспоненциальному закону общее количество информации [1], ужесточение требований по ее хранению, поиску и обработке, увеличение трафика и скорости передачи информации предопределили появление информационных систем (ИС) различных поколений и назначения. Сегодня термин ИС охватывает автоматизированные системы, компьютерные сети или системы связи [2], информационно-телекоммуникационные системы [3] и т.д. В ИС концентрируется и циркулирует большой объем как открытой, так и информации с ограниченным доступом (ИсОД).

В связи с этим приобрела широкий размах и деятельность по гласному и негласному добыванию информации из открытых и закрытых ИС, баз и банков данных, контролю за сообщениями, передаваемыми в вычислительных сетях, получению персональных данных пользователей ИС и другой ценной компьютерной информации. Для характеристики подобной деятельности стали широко использоваться термины: «компьютерный шпионаж», «компьютерная разведка», «информационно-аналитическая работа в Интернет», «аналитическая разведка», «компьютерный анализ и разведка» и др.

Однако в нормативно-методических документах и многочисленных публикациях по данной тематике до сих пор отсутствует единое терминологическое толкование сущности, задач и методов компьютерной разведки (КР), что и обуславливает актуальность рассматриваемой проблемы.

Ряд авторов, специализирующихся на теории и практике экономической разведки (называемой также конкурентной, деловой, коммерческой, competitive intelligence, business intelligence и др.), определяют КР как аналитическую обработку огромного числа данных из разнообразных открытых источников информации, прежде всего из Интернет. Сущность КР они видят в поиске и передаче информации из открытых компьютерных систем и сетей “всемирной паутины” с последующей верификацией и аналитической обработкой [4-5].