УДК 621.391.28

МОДЕЛИ АНАЛИЗА ДАННЫХ МОНИТОРИНГА АВТОМАТИЗИРОВАННЫХ СИСТЕМ

Н.Н. Мошак, доцент С.-Петербургского государственного университета телекоммуникаций (СПбГУТ) им. проф. М.А.Бонч-Бруевича
Е.А. Тимофеев, аспирант СПбГУТ им. проф. М.А.Бонч-Бруевича

Мониторинг, или контроль состояния автоматизированных систем (АС) является одной из основных функций ее подсистемы информационной безопасности (ИБ), которая реализует требования политики ИБ АС [1]. Основной задачей мониторинга является наблюдение в режиме реального времени за активными субъектами системы (пользователями, процессами и др.), сбор и анализ данных о функционировании этих субъектов, обобщение результатов анализа и формирование качественной оценки функционирования системы в целом.

Итоговая оценка может содержать информацию о доступности информационных активов системы, об обнаружении деструктивных воздействий и др. Результат указанной оценки оформляется в виде сообщений «несанкционированное действие» (НСД) и/или «инцидент» (событие, требующее расследования) и пересылается администратору АС.

В подсистемах мониторинга может применяться как сигнатурный, так и статистический анализ данных. Под сигнатурным анализом понимается анализ событий мониторинга, выделяемых в потоке регистрируемых данных, на соответствие заданным событийным цепочкам (сигнатурам), определяемым специально настраиваемыми правилами (шаблонами, фильтрами). Каждое из правил представляет собой цепочку первичных событий с определенными параметрами (контекстом события). Анализируемый поток первичных событий анализируется на предмет присутствия в нем заданных цепочек. Сигнатурный анализ производится в режиме реального времени. С правилами связаны обработчики событий, в которых производится анализ контекста на соответствие данному правилу и вырабатывается необходимая реакция, а именно: генерируется результирующее событие, сообщение о НСД или инциденте и т. д. Другими словами, методы сигнатурного анализа позволяют фиксировать факты конкретных нарушений, механизм которых заранее известен.

Статистический анализ дает возможность оценить соответствие текущего сеанса поведения легальных субъектов АС шаблону их поведения, построенному за определенные периоды времени. При статистической обработке анализируются все события, фиксируемые системой мониторинга, а не только сигнатурные цепочки заданных событий. При этом предполагается, что существуют статистические параметры наблюдаемой системы, неизменность которых отражает стабильность характера поведения ее субъектов. Количественные изменения этих параметров свидетельствует либо об изменениях поведения субъектов (в том числе связанных с несанкционированными действиями), либо о действиях субъектов, направленных на дестабилизацию системы (изучение ее устойчивости, исследование недокументированных возможностей, обнаружение ошибок в администрировании и т. д.). Необходимо отметить, что типичной ситуацией в АС является невозможность с помощью правил безопасности ограничить нежелательные действия легального пользователя, не лишив его при этом необходимой функциональности. В этой связи применение статистических методов играет большую эвристическую роль в системе мониторинга АС, позволяя, во-первых, выявлять нетипичные действия легальных субъектов и, во-вторых, проводить более детальный анализ и дополнительные расследования.

Модель поведения субъекта АС. В большинстве случаев система мониторинга строится как независимый элемент АС, т. е. как элемент, неподконтрольный ее административному персоналу и управляемый только администратором ИБ АС. При этом данные мониторинга поступают в систему как от специальных программных агентов, устанавливаемых на автоматизированных рабочих местах (АРМ) пользователей и на серверах АС, так и от штатных систем регистрации событий элементов АС (операционных систем, СУБД, маршрутизаторов и др.) [3]. Необходимо отметить, что при рассмотрении статистики наблюдаемых параметров обычно недопустимо предполагать их независимость [4, 5]. В частности, это так, если в качестве параметров выбираются их аддитивные величины, которые представляют собой отдельные суммы «элементарных» событий различного типа (например, вызов некоторой функции API Windows, длительность сеанса, и др.), накапливаемых в течение сеанса работы субъекта [6]. Как известно, статистический корреляционный анализ [7] не дает объяснения наблюдаемым зависимостям между параметрами. Однако предварительный анализ наблюдаемых параметров часто помогает предсказать наличие статистических связей между ними и их характер [4, 5].

Вектор $\vec{X}(x_1, x_2 \ldots, x_n)$ в n-мерном пространстве наблюдаемых параметров, соответствующий некоторому сеансу, будем называть *вектором сеанса*. Вектор $\vec{e^j}(e_1^j, e_2^j \ldots, e_n^j), j = \overline{1,k}$, соответствующий j-й «элементарной» операции или циклу, будем в дальнейшем называть *циклом*. При этом, если векторы сеансов принадлежат линейному k-мерному подпространству n-мерного пространства наблюдаемых параметров, то наблюдаемые параметры сеансов связаны линейной зависимостью.

Принадлежность векторов сеансов линейному подпространству можно объяснить «векторной» моделью поведения субъекта, которая строится следующим образом. Предположим вначале, что функциональный профиль или поведение субъекта в АС определяется циклическим повторением одной «элементарной» операции. Тогда сеансы работы субъекта будут однозначно определяться количеством выполненных циклов. Значения каждого из наблюдаемых параметров, подсчитанные за сеанс, будут пропорциональны количеству выполненных циклов в данном сеансе. В этом случае точки $\vec{X}$ в n-мерном пространстве параметров, соответствующие различным сеансам, располагаются на одной прямой, ориентация которой определяется вектором $\vec{e}^1(e_1^1, e_2^1 \ldots, e_n^1)$, соответствующим одному циклу операций.

Усложним поведение контролируемого субъекта: допустим, что субъект может выполнять две «элементарные» операции. В этом случае появляется второй цикл, и точки, соответствующие различным сеансам, будут заполнять плоскость, определенную векторами $\vec{e}^1$ и $\vec{e}^2$ (более точно: угол этой плоскости, заполняемый векторами, являющимися сум-

мой исходных циклов, умноженных на положительные коэффициенты). Таким образом, аппроксимируя распределение некоторой выборки сеансов линейным подпространством, которое в дальнейшем будем называть пространством сеансов — соответствующей размерности, и, оценивая близость текущего сеанса к аппроксимирующему пространству, можно контролировать соответствие текущего сеанса статистическому характеру (шаблону) поведения субъекта.

Разложение вектора сеансов $\vec{X}$ по выполненным k циклам операций имеет вид

$$X = m^1\vec{e}^1 + m^2\vec{e}^2 + ... + m^k\vec{e}^k + \vec{\delta}.$$

Здесь $m^j, j = \overline{1,k}$ — количество выполненных за сеанс циклов j-го типа, $\vec{e}^j$ — базисный вектор, соответствующий циклу i-го типа, $\vec{\delta}$ — погрешность, определяющая разброс за счет неучтенных циклов. Количество циклов (размерность пространства) сеансов может быть меньше или равным размерности пространства наблюдаемых параметров. Увеличивая количество наблюдаемых параметров (или ограничивая количество допустимых циклов), можно всегда достичь ситуации $k \lhd n$, т. е. когда пространство сеансов будет некоторым подпространством наблюдаемых параметров. Неучтенные, редко выполняемые операции приводят к статистическому разбросу точек сеанса относительно подпространства сеансов модели поведения субъекта.

Если размерность пространства сеансов меньше количества выполняемых циклов, то мы имеем вырожденный случай: циклам операций соответствуют линейно-зависимые векторы $\vec{e}^j$ в пространстве параметров. Количество линейно независимых циклов определяет количество «степени свободы» контролируемой системы. Всегда можно ограничиться некоторым количеством наиболее весомых (степень весомости определена ниже) «степеней свободы», определяющих поведение субъектов АС, а влияние оставшихся — отнести к статистическому разбросу. На основании вышеизложенного модель поведения субъекта АС можно представить в виде линейного пространства сеансов. Ниже с учетом этой модели будет построена статистическая модель анализа данных АС, формализующая оценку статистического разброса вектора сеанса X от линейного пространства сеансов типичного поведения субъекта или типичного функционального профиля его работы. Указанная оценка характеризует состояние системы в целом.

Построение аппроксимирующего пространства. Критерий применимости линейной аппроксимации. Аппроксимацию статистической линейной зависимости наблюдаемых параметров пространства сеансов проведем на базе метода максимального правдоподобия. При этом пространство сеансов будет некоторым гиперпространством (размерности $n - 1$). Пространства меньшей размерности являются частным случаем полученных результатов. Этот подход используется при оценке экспериментальных данных, в предположении, что истинные значения наблюдаемых параметров связаны линейной зависимостью, а отклонения измеренных значений от истинных значений распределены по нормальному закону. Ставится задача нахождения параметров, связанных линейной зависимостью, для которых минимальна дисперсия отклонений точек сеансов конкретной выборки от аппроксимирующего пространства [7]. Для определения меры (т. е. расстояния между точками углов, от точки до плоскости и т. д.) в n-мерном пространстве наблюдаемых параметров они обычно предварительно нормируются на свою дисперсию.

Задача аппроксимации гиперпространством соответствующей размерности пространства сеансов в пространстве наблюдаемых параметров с учетом их линейной зависимости методом максимального правдоподобия сводится к решению задачи на собственные значения для матрицы ковариации R. Аппроксимирующее гиперпространство в общем виде задается уравнением:

$$\sum_{j=1}^{n} \alpha_j X_j = C \qquad (1)$$

(для гиперпространства, проходящего через начало координат $C = 0$). Данное уравнение определяет условие постоянства (равенства нулю) скалярного произведения радиус-вектора, соответствующего точкам гиперпространства $\{X_j\}$ на некоторый вектор $\{\alpha_j\}$, определяющий это гиперпространство. Очевидно, что данное условие однозначно определяет гиперпространство или линейную зависимость переменных $X_1, X_2, ... X_n$. Коэффициенты α_i варьируются для поиска такого вектора $\{\alpha_i\}$, для которого минимизируется сумма квадратов расстояний от точек сеансов конкретной выборки размера M сеансов ξ_j^i работы субъекта АС ($i = \overline{1, M}$ — номер сеанса) до аппроксимирующего гиперпространства. Все переменные подвержены случайным ошибкам, так что точки сеансов конкретной выборки имеют вид:

$$\xi_j^i = X_j^i + \delta_j^i, \text{ где} \qquad (2)$$

$j = 1, 2, ..., n$ для всех i. Предполагается, что случайные ошибки δ_j^i нормально распределены и не зависят от X_j^i, друг от друга и имеют нулевые средние. В этом случае логарифм функции правдоподобия L имеет вид [7]:

$$\log L = const - nM \log \sigma_\delta - \frac{1}{2\delta^2} \sum_{j=1}^{n} \sum_{i=1}^{M} (\xi_j^i - X_j^i)^2. \qquad (3)$$

Если рассматривать нашу выборку, как M точек в n-мерном пространстве параметров, то поставленная задача будет заключаться в определении гиперплоскости (1). Максимизация функции правдоподобия L эквивалентна минимизации двойной суммы в выражении (3). Учитывая, что расстояние D от точки $(\xi_1^i, ..., \xi_n^i)$ до гиперплоскости (1) дается выражением

$$\Delta = \frac{\sum_{j=0}^{n} \alpha_j \xi_j^i}{(\sum_{j=0}^{n} \alpha_j^2)^{\frac{1}{2}}}, \qquad (4)$$

то сформулированная задача сводится к минимизации функции:

$$S = \frac{\sum_{i=1}^{M} (\sum_{j=0}^{n} \alpha_j \xi_j^i)^2}{\sum_{j=0}^{n} \alpha_j^2}. \qquad (5)$$

При варьировании коэффициентов α_i достаточно изменять только направление вектора $\{\alpha_i\}$, поскольку изменение его модуля не приводит к изменению гиперпространства. Другими словами, задача (5) может быть сформулирована на условный экстремум следующим образом: минимизировать сумму $S' = \sum_i (\sum_j \alpha_j \xi_j^i)^2$ при ограничении $\sum \alpha^2 = \text{const}$.

С помощью неопределенного множителя Лагранжа μ эта задача сводится к нахождению безусловного минимума

$$\min_{\alpha_l} \sum_i (\sum_j \alpha_j \xi_j^i)^2 - \mu \sum_j \alpha_j^2. \quad (6)$$

После дифференцирования по α_l, получаем:

$$\sum_i \xi_l^i (\sum_j \alpha_j \xi_j^i) = \mu \alpha_l. \quad (7)$$

Характеристическое уравнение для нахождения собственных значений матрицы ковариации R (с учетом сдвига начала координат и нормировки на среднее квадратичное отклонение матрица ковариации совпадает с матрицей корреляции) имеет следующий вид:

$$\begin{vmatrix} 1-\theta_1 & r_{12} & r_{13} & r_{1n} \\ r_{12} & 1-\theta_2 & r_{23} & r_{2n} \\ r_{13} & r_{23} & 1-\theta_3 & r_{3n} \\ r_{1n} & r_{2n} & r_{3n} & 1-\theta_n \end{vmatrix} = 0, \quad (8)$$

где $\theta_i = \dfrac{\mu}{M S_i^2}$, $r_{ij} = (\sum_{l=1}^{M} \xi_i^l \xi_j^l)/M$ — элементы матрицы ковариации R.

Определенная таким образом матрица R является инвариантной относительно ортогональных преобразований (поворотов в R^n) величиной:

$$r'_{ij} = M^{-1} \sum_{l=1}^{M} \xi'_{il} \xi'_{jl} = M^{-1} \sum_{l=1}^{M} \sum_{p=1}^{n} U_{ip} \xi_{pl} \sum_{m=1}^{n} U_{jm} \xi_{ml} =$$

$$= M^{-1} \sum_{l=1}^{M} \sum_{p=1}^{n} U_{ip} \xi_{pl} \sum_{m=1}^{n} \xi_{ml} U^{-1}{}_{jm} = \sum_{p=1}^{n} U_{ip} \sum_{m=1}^{n} M^{-1} (\sum_{l=1}^{M} \xi_{pl} \xi_{ml}) U^{-1}{}_{jm} =$$

$$= \sum_{p=1}^{n} U_{ip} \sum_{m=1}^{n} r_{pm} U^{-1}{}_{jm}.$$

Таким образом, $R' = URU^{-1}$, где R и R' — матрицы ковариаций, вычисленные в различных базисах (т. е. матрица ковариации относительно ортогональных преобразований ведет себя так же, как линейные операторы в R^n). При нахождении безусловного минимума нас интересует минимальный корень $\mu_{\min}$ уравнения (8), поскольку в точках экстремума $S = \mu \sum \alpha_i^2$. Соответствующий собственный вектор задает гиперпространство, аппроксимирующее пространство сеансов в соответствии с (1) и ортогонален этому гиперпространству.

Таким образом, критерием применимости аппроксимации пространства сеансов является то, что одно или несколько собственных значений матрицы R существенно меньше, чем остальные минимальные значения одного или нескольких собственных значений $\mu_{\min}$ матрицы R (относительно собственных значений $\mu_i, i = \overline{1,n}$). Метод максимального правдоподобия при этом позволяет автоматически учесть размерность пространства сеансов.

Можно представить наглядную геометрическую интерпретацию этого факта. Для чего введем понятие эллипсоида рассеяния, который определяется постоянством на его поверхности квадратичной формы:

$$f^2 = \sum_{i,k=1}^{n} a_{ik} x_i x_k, \quad (9)$$

и следующим дополнительным условием: если плотность вероятности точек сеансов «нормального» или априорного поведения субъекта равна постоянному значению внутри этого эллипсоида и нулю вне него, то априорное распределение дает такие же моменты второго порядка (матрицу ковариации R), что и рассматриваемое распределение конкретной выборки. Другими словами, эллипсоид рассеяния моделирует наблюдаемое распределение случайной величины (точки сеанса) в приближении равномерного распределения результатов наблюдения в пределах данного эллипсоида, сохраняя при этом моменты второго порядка реального распределения.

Моменты второго порядка эллипсоида (9) равны $\dfrac{f^2}{n+2} \dfrac{a_{ik}}{|a|}$, отсюда $f^2 = n+2$, а $\dfrac{a_{ik}}{|a|} = r_{ik}$ — элементы матрицы ковариаций R наблюдаемых параметров. Объем эллипсоида рассеяния V с учетом принятой нормировки дается выражением

$$V = \frac{(n+2)^{n/2} \pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} \sqrt{|R|},$$

где $\Gamma\left(\dfrac{n}{2}+1\right)$ — гамма-функция по аргументу n. В случае независимости наблюдаемых параметров недиагональные элементы матрицы R равны нулю, определитель $|R| = 1$ и объем эллипсоида рассеяния максимален. Чем больше эллипсоид рассеяния сосредоточен в окрестности некоторого линейного подпространства (его некоторые полуоси существенно меньше других), тем меньше объем эллипсоида рассеяния и меньше коэффициент рассеяния $\sqrt{|R|}$. Коэффициент рассеяния является характеристикой некоторой конкретной статистической выборки и может использоваться как критерий существования линейной зависимости и применимости какого-либо метода линейной аппроксимации или правильного выбора наблюдаемых параметров (например, метода регрессии или метода максимального правдоподобия).

Модель статистического анализа данных мониторинга АС. В самом общем виде задача по статистической оценке состояния АС может быть поставлена следующим образом: есть некоторый вектор X_j наблюдаемых параметров поведения субъекта (текущий сеанс работы субъекта), требуется оценить его значение на предмет соответствия ранее построенному априорному функциональному профилю за определенные периоды времени. Если имеется ряд априорных состояний системы, определяющих набор классов, каждый из которых характеризуется своим распределением сеансов, то классификация текущего состояния является типовой задачей распознавания образов. Методы теории распознавания образов позволяют отнести отдельный сеанс к одному из классов (классифицировать) и дают ошибку классификации в соответствии с одним из критериев (например, Байеса, Неймана-Пирсона, минимакса и др.)

Однако существует ряд задач, которые рассматривают проблему определения состояния системы как «норма» или «отклонение». При этом известное распределение имеется только для «нормального» ее состояния. В случае статистической линейной зависимости наблюдаемых параметров распределение сеансов (определяющее «норму» поведения) сосредоточено в окрестности определенного выше аппроксимирующего пространства. Плотность распределения максимальна в точках аппроксимирующего пространства и убывает при удалении от нее. В качестве меры скорости убывания плотности распределения будем использовать среднеквадратичное отклонение

точек сеанса от аппроксимирующего пространства. Поэтому в основу статистической модели анализа данных мониторинга АС могут быть положены два критерия: «стабильность», который характеризует сосредоточенность распределения вокруг аппроксимирующего пространства, и «нормальность», характеризующий отклонение наблюдаемого сеанса от аппроксимирующего пространства в сравнении с среднеквадратичным отклонением. Критерий «стабильность» может быть задан как условие малости среднеквадратичного отклонения точек сеансов ξ_j^i конкретной выборки M от аппроксимирующего пространства по сравнению с разбросом точек внутри него. Из выводов предыдущего параграфа ясно, что этот критерий может быть формализован в виде

$$\sqrt{\mu_{\min}} \lhd\lhd \sqrt{\mu} ,$$

где μ — собственные значения матрицы ковариации, за исключением одного или нескольких минимальных. Эквивалентная формулировка критерия «стабильность» — малое значение (по сравнению с единицей) коэффициента рассеяния $\sqrt{|R|}$. Критерий «нормальность» относится к отдельному наблюдаемому сеансу X_j работы субъекта. После преобразования координат в пространстве наблюдаемых параметров к базису собственных векторов матрицы ковариации отклонение наблюдаемого сеанса от аппроксимирующего пространства (в сравнении со среднеквадратичным отклонением) задается координатой, соответствующей $\mu_{\min}$, поэтому критерий можно формализовать в виде

$$X_j'^i < \sqrt{\mu_{\min}} .$$

Для оценки «отклонений» метрика определяется нормировкой всех параметров на их среднее квадратичное отклонение (из предыдущего параграфа нетрудно понять, что для введенных критериев это не принципиально). Кроме того, начало координат для удобства выбирается в точке, координаты которой соответствуют математическим ожиданиям наблюдаемых параметров. Распределение сеансов выборки ξ_j^i в этом случае сосредоточено в окрестности куба со сторонами [-1,1]. На рис. 1 представлен вид распределений точек сеансов ξ_j^i для примера двух наблюдаемых параметров в зависимости от их коррелированности в виде эллипсоидов рассеяния.

Как известно, симметричные вещественные матрицы (симметричные операторы в R^n) имеют n вещественных собственных значений. Соответствующие собственные векторы μ_j взаимно-ортогональны и образуют базис (в n-мерном пространстве наблюдаемых параметров). Если произвести преобразование системы координат к собственным векторам матрицы корреляции, то она примет диагональный вид. При этом «фиктивные» наблюдаемые параметры некоррелированы, собственные значения равны квадрату среднего квадратичного отклонения. Минимальное собственное значение $\mu_{\min}$ определяет среднеквадратичное отклонение точек сеансов $X_l'^i$ от аппроксимирующего гиперпространства. «Фиктивный» наблюдаемый параметр, соответствующий этому собственному значению X_l', дает отклонение точки текущего сеанса от аппроксимирующего гиперпространства. Линейная оболочка собственных векторов матрицы ковариации R, за исключением собственного вектора, соответствующего минимальному собственному значению $\mu_{\min}$, представляет собой гиперпространство, аппроксимирующее статистическую линейную зависимость наблюдаемых параметров, или пространство сеансов, введенное выше.

Предложенные критерии статистической оценки состояния АС были опробованы на реальных данных для автоматизированной банковской системы (АБС). При этом для большинства АРМ оказалось достаточным совместное рассмотрение 3—4 наблюдаемых параметров (коэффициент рассеяния при этом < 0,01). Таким образом, пользователь данной АБС, как правило, выполняет 2—3 основные операции. Кроме того, экспериментально подтверждена возможность выявления «нетипичных» сеансов по критерию «нормальность» для данной модели. С этой целью к выборке сеансов некоторого пользователя добавлялись данные, относящиеся к сеансу другого пользователя. При этом значение критерия «нормальность» поведения субъекта в сеансе существенно отличалось от остальных.

Для трехмерной реализации алгоритма аппроксимации (трех наблюдаемых параметров) в системе MathCad строились распределения в трехмерном пространстве наблюдаемых параметров и соответствующие аппроксимирующие плоскости. В этом случае возможна объективная визуальная оценка близости точек сеансов наблюдения к аппроксимирующей плоскости, благодаря возможности вращения системы координат для 3D-графиков. Результат выборочных просмотров статистических данных показал, что практически для любого субъекта АБС существуют три параметра, связь которых достаточно стабильна, т. е. хорошо аппроксимируется плоскостью. Для наглядности производится преобразование ко-

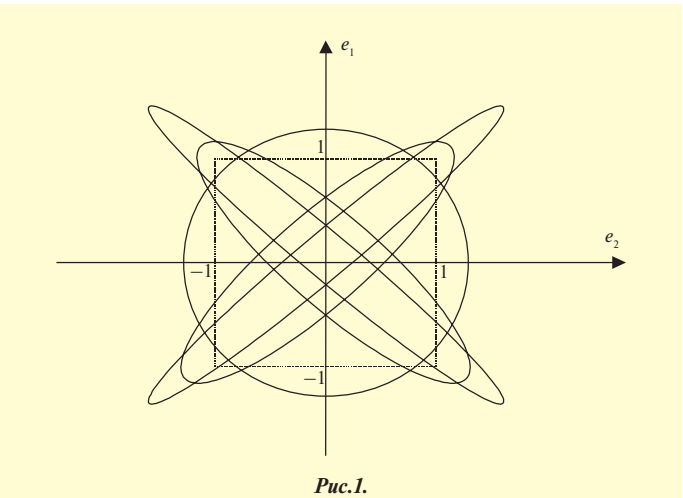

Рис.1.

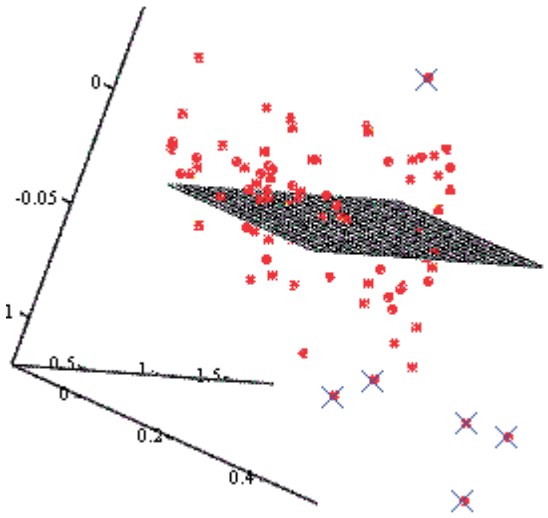

Рис. 2.

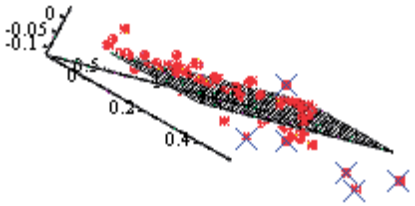

Рис. 3.

ординат в пространстве наблюдаемых параметров, такое, что одна из координатных осей становится ортогональна аппроксимирующей плоскости. Тогда значение соответствующей координаты текущего сеанса входит в критерий «нормальность» данного сеанса, а среднее квадратичное отклонение этой координаты по выборке сеансов — в критерий «стабильность».

На рис. 2 масштаб растянут по оси, перпендикулярной аппроксимирующей плоскости, а на рис. 3 включена опция “Equal Scales” просмотра графиков 3D. Голубыми крестами отмечены «подозрительные» сеансы, т. е. отклонение которых от плоскости больше заданной величины (0,05).

В случае размерности, большей трех, преобразование координат — единственная возможность визуально оценить, насколько стабильна связь между величинами. В качестве примера приведена серия двухмерных графиков, построенных с использованием этого метода — проекций трехмерного гиперпространства, аппроксимирующего взаимозависимость четырех наблюдаемых параметров для той же выборки. Проекции берутся на координатные плоскости в системе координат, одна из осей которой ортогональна аппроксимирующей плоскости. Дельта для определения подозрительных сеансов уменьшена до 0,04, поскольку для дельты, равной 0,05, «подозрительных сеансов» нет.

На рис. 4 изображены проекции на остальные пары координатных плоскостей, на координатные оси которых натянуто аппроксимирующее гиперпространство. На рис. 5 изображены проекции на координатные плоскости, для которых одна из осей совпадает с перпендикуляром к аппроксимирующему трехмерному гиперпространству.

Данный подход может быть, в принципе, распространен на произвольную размерность пространства выборки. При этом оценка параметра «стабильность» для данной выборки и «нормальность» текущего сеанса может производиться не визуально, а по значениям соответствующей координаты.

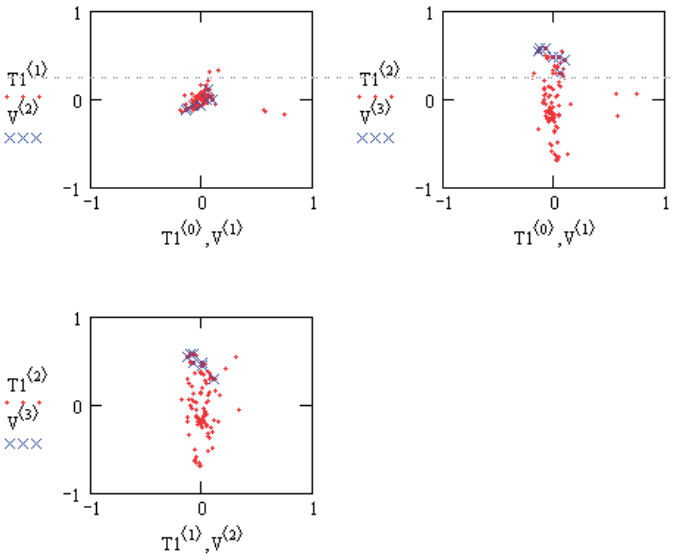

Рис. 4

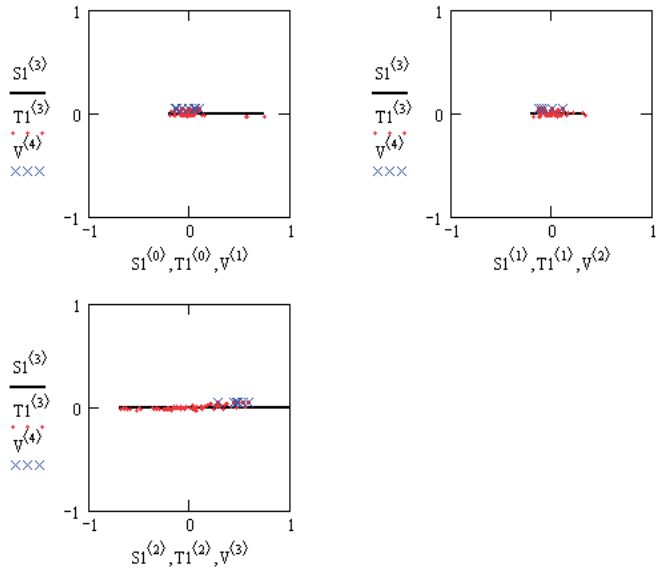

Рис. 5

Результаты обработки данных мониторинга АБС сведены в таблицу. Аппроксимация линейной зависимости наблюдаемых параметров проводилась дополнительно с исполь-

Таблица

Подконтрольный пользователь	Размер выборки	Наблюдаемые параметры	Коэффициент рассеяния	$\sigma = \sqrt{\mu}$	Количество “подозрительных сеансов” n для допуска δ	
					n	δ
ODB_102	49	W1, W2, F1	0,058	0,095	7	0,15
		W2, F1, F2	0,019	0,043	5	0,07
		W2,U1, U2	0,00045	0,00618	6	0,01
ODBV_701	15	W1, W2, F1	0,864	0,797		
		W1, W2, F2	0,719	0,541		
ODB_103	52	W1, W2, F2	0,007973	0,011	1	0,003
					4	0,02
ODB_12	52	W1, W2, F1	0,081	0,137	8	0,2
		W2, F1, F2	0,132	0,203	2	0,4

зованием метода регрессии. В таблице для подконтрольных субъектов (пользователей) на основании приведенных выборок приведены значения коэффициента рассеяния, среднеквадратичного отклонения от плоскости аппроксимации, определяющего критерий «стабильность» ($\mu_{min}^2 = \sigma$), а также количество сеансов, не удовлетворяющих критерию «нормальность» при заданном допустимом отклонении δ).

Итак:

- имеются явно выраженные линейные зависимости наблюдаемых параметров (малый коэффициент рассеяния);
- наблюдается сосредоточенность распределения около аппроксимирующей плоскости (малое значение σ).

Для любого пользователя можно найти комбинации наблюдаемых параметров, для которых линейное гиперпространство подходящим образом аппроксимирует их взаимную зависимость.

Выводы. 1. **Статистическую линейную зависимость наблюдаемых параметров сеансов работы субъекта АС предлагается аппроксимировать линейным пространством.** Критерием применимости такой аппроксимации является малость одного или нескольких собственных значений матрицы корреляций наблюдаемых параметров (относительно остальных собственных значений).

2. **Предложенная модель статистического анализа данных мониторинга АС в виде критерия «нормальность» сеанса работы субъекта АС, при условии выполнения критерия «стабильность» для исследуемой выборки сеансов, позволяет** оценить нештатное поведение субъектов АС, в отличие от моделей сигнатурного анализа, дающих возможность выявить только нарушение конкретных требований.

ЛИТЕРАТУРА

1. **Мошак Н.Н., Тимофеев Е.А.** Особенности построения политики информационной безопасности в инфокоммуникационной сети // Электросвязь. — 2005. — № 9.
2. СТО БР ИББС-1.0-2006 «Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Общие положения».
3. **Галатенко А.В.** Активный аудит// JetInfo. — 1999. — № 8.
4. **Тимофеев Е.А.** Использование статистических методов в системах мониторинга информационной безопасности автоматизированных систем Банка России// Вестник Северо-Запада. Информационно-аналитический бюллетень. — 2006. — № 4(35).
5. **Тимофеев Е.А.** Модели обработки данных мониторинга информационной безопасности в автоматизированных банковских системах. В сб. «VI международная научная конференция «Информационные сети, системы и технологии». — СПб. — 2006.
6. Система наблюдения и контроля за работой клиентов в платежных системах, организованных по принципу «клиент-сервер». Общее описание системы». МПИФ.00012-01 94-ЛУ.
7. **Кендалл М.Дж., Стюарт А.** Статистические выводы и связи. Т. 2. — М.: Наука, 1973.
8. **Фукунага Г.** Введение в теорию распознавания образов. — М.: Наука, 1972.

Получено 3.03.08