Состоятельность долгосрочных профилей в системах выявления вторжений

М. А. Ермохин (МГУ им. М. В. Ломоносова, Москва)

В работе исследуется модель долгосрочных профилей систем активного аудита, предложенная А. В. Галатенко, А. Е. Лебедевым и И. Н. Емельяновым в 2009 году. Доказывается, что предложенное авторами модели достаточное условие сходимости профилей является критерием.

Ключевые слова: выявление вторжений, состоятельность долгосрочных профилей.

Введение

В настоящее время системы активного аудита, также называемые системами выявления вторжений или системами предупреждения вторжений, играют важную роль в обеспечении компьютерной безопасности. Содержательно говоря, задачей таких систем является выявление злоумышленной деятельности (как правило, задаваемой экспертами в виде набора регулярных выражений) и нетипичной активности. Большинство решений ограничивается выявлением злоумышленных действий. Заметим, что при этом, с одной стороны, может возникнуть проблема экспоненциального взрыва числа состояний распознающего автомата (см., например, [1]), и, с другой стороны, невозможно выявление атак, не описанных экспертами.

Один из возможных способ выявления неизвестных атак заключается в выделении ситуаций, являющихся в том или ином смысле нетипичными; нетипичность может быть следствием проводимой или успешно проведенной атаки или программного или аппаратного сбоя. Задача выявления нетипичного поведения может быть сведена к двум поздадачам: описанию типичного поведения (долгосрочных профилей) и сравнению текущего поведения с типичным. Подробнее остановимся на решении первой подзадачи. Отметим, что более подробную информацию о системах активного аудита можно найти в работе [2].

Авторы систем выявления нетипичной активности явно указывают, что типичное поведение может изменяться с течением времени (см., например, [3]). При этом желательно, чтобы более поздние события учитывались с большим весом. Для достижения этой цели предлагалось либо учитывать только события за промежуток времени фиксированной длины, «забывая» предысторию (так называемое скользящее окно), либо проводить усреднение с экспоненциально убывающими весами ([3]). При этом в качестве обоснования подхода приводились либо эвристические соображения, либо результаты эксперимента. В работе [4] была введена математическая модель, для которой и метод скользящего окна, и метод усреднения с убывающими весами оказывались частными случаями. Там, в частности, исследовался вопрос состоятельности профилей, то есть сходимости по вероятности к математическому ожиданию. Было предложено достаточное условие состоятельности, сформулированное в терминах требований к параметрам модели, и показана необходимость условия для ряда частных случаев. Мы докажем, что достаточное условие на самом деле является критерием состоятельности.

Основные понятия и результаты

Общая схема построения долгосрочных профилей может быть описана с помощью следующей модели, введенной в работе [4]: пусть $\{x_t\}_{t=1}^{\infty}$ — последовательность независимых дискретных случайных величин, принимающих значения из множества $\{1, \ldots, M\}$. Обозначим $\mathbb{P}^l = \{(t_1, \ldots, t_l) \in \mathbb{R}^l \mid \forall i \in \overline{1,l}\ t_i > 0, \sum\limits_{i=1}^{l} t_i = 1\}$ Предположим, что задано семейство M-мерных векторов

$$P(t) = (p_1(t), \ldots, p_M(t)) \in \mathbb{P}^M,$$

и для каждого t случайная величина x_t имеет распределение, задаваемое вектором $P(t)$, то есть $\forall t \in \mathbb{N}\ \forall m \in \overline{1,M}\ \mathtt{P}\{x_t = m\} = p_m(t)$. Обозначим также $q_m(t) = 1 - p_m(t)$.

Положим $\tilde{p}_m(t) = \sum\limits_{\tau=1}^{t} \omega_\tau(t) I_{\{x_\tau = m\}}$, где $\omega(t) = (\omega_1(t), \ldots, \omega_t(t))$ — некоторый t-мерный весовой вектор ($\omega_\tau(t) \geqslant 0$, $\tau = 1, \ldots, t$), заданный для каждого момента времени t. Всюду далее мы будем считать, что конечные суммы весов равномерно ограничены константой $B > 0$ и отделены от нуля, то есть

$$0 < \sum_{\tau=1}^{t} \omega_\tau(t) < B.$$

Вектор $\tilde{P}(t) = (\tilde{p}_1(t), \ldots, \tilde{p}_M(t))$ будем называть вектором частот. Обратим внимание, что случай $\omega_\tau(t) = ab^{-c(t-\tau)}$, где a, b и c — неотрицательные константы, задает усреднение с экспоненциально убывающими весами, а случай $\omega(t) = \Big(\underbrace{0, \ldots, 0}_{t-n}, \underbrace{\frac{1}{n}, \ldots, \frac{1}{n}}_{n}\Big)$ соответствует схеме «скользящее окно».

Нас интересует, насколько вектор частот отличается от вектора истинного распределения.

Теорема 1. *Для того, чтобы* $\forall \varepsilon > 0 \lim\limits_{t\to\infty} \mathtt{P}\{|\tilde{p}_m(t) - \mathtt{E}\tilde{p}_m(t)| \geqslant \varepsilon\} = 0$, *необходимо и достаточно, чтобы* $\lim\limits_{t\to\infty} \sum\limits_{\tau=1}^{t} \omega_\tau^2(t) p_m(\tau) q_m(\tau) = 0$.

Доказательство теоремы. *Необходимость.* В силу независимости x_t, $\sum\limits_{\tau=1}^{t} \omega_\tau^2(t) q_m(\tau) p_m(\tau) = \mathtt{D}\tilde{p}_m(t)$. Таким образом, необходимость является прямым следствием следующей леммы.

Лемма 1. *Пусть последовательность дискретных случайных величин* $\{\sigma_t\}_{t=1}^{\infty}$ *равномерно ограничена константой* $B > 0$, *неотрицательна и сходится по вероятности к константе* C. *Тогда дисперсия величины* $(\sigma_t - C)$ *стремится к* 0 *при* $t \to \infty$.

Доказательство леммы. Пусть значения σ_t — множество $A_t = \{a_{1t}, \ldots, a_{n_t t}\}$ Необходимо для произвольного $\varepsilon > 0$ найти такое $T \in \mathbb{N}$, что $\forall t > T$ $\mathtt{D}\sigma_t < \varepsilon$. Рассмотрим $\varepsilon' > 0$ и $\delta' > 0$, значения которых будут выбраны позже. По определению дисперсии имеем:

$$\mathtt{D}(\sigma_t - C) = \mathtt{E}(\sigma_t - \mathtt{E}\sigma_t)^2 = \\ = \sum_{\tau=1}^{n_t} (a_{\tau t} - \mathtt{E}\sigma_t)^2 \mathtt{P}\{\sigma_t = a_{\tau t}\} = D_1 + D_2,$$

где

$$D_1 = \sum_{\tau : |a_{\tau t} - C| \geqslant \varepsilon'} (a_{\tau t} - \mathtt{E}\sigma_t)^2 \mathtt{P}\{\sigma_t = a_{\tau t}\},$$

$$D_2 = \sum_{\tau : |a_{\tau t} - C| < \varepsilon'} (a_{\tau t} - \mathtt{E}\sigma_t)^2 \mathtt{P}\{\sigma_t = a_{\tau t}\}.$$

Оценим D_1. В силу сходимости по вероятности величины $\sigma_t - C$ к 0 найдется $T' \in \mathbb{N}$, такое что для всех $t > T'$ $\mathtt{P}\{|\sigma_t - C| \geqslant \varepsilon'\} < \delta'$. Так как, очевидным образом $|\sigma_t - \mathtt{E}\sigma_t| < 2B$, для любого $t > T'$ справедливы следующие неравенства:

$$D_1 \leqslant \sum_{\tau:|a_{\tau t}-C|\geqslant\varepsilon'} 4B^2\mathtt{P}\{\sigma_t = a_{\tau t}\} = 4B^2 \sum_{\tau:|a_{\tau t}-C|\geqslant\varepsilon'} \mathtt{P}\{\sigma_t = a_{\tau t}\} < 4\delta' B^2.$$

Оценим D_2. Выполняются следующие неравенства:

$$D_2 < \sum_{\tau:|a_{\tau t}-C|<\varepsilon'} (\varepsilon')^2\mathtt{P}\{\sigma_t = a_{\tau t}\} < (\varepsilon')^2 \sum_{\tau:|a_{\tau t}-C|<\varepsilon'} \mathtt{P}\{\sigma_t = a_{\tau t}\} < (\varepsilon')^2.$$

Таким образом, для всех $t > T'$ справедлива оценка $\mathtt{D}(\sigma_t - C) < 4\delta' B^2 + (\varepsilon')^2$. Пусть $(\varepsilon')^2 < \varepsilon$ и $\delta' < \frac{\varepsilon-(\varepsilon')^2}{4B^2}$. Тогда при $T' = T$ и любом $t > T$ имеем $\mathtt{D}(\sigma_t - C) < \varepsilon$. Лемма доказана.

Замечание. Условие равномерной ограниченности является существенным. В качестве примера достаточно рассмотреть последовательность σ_t, принимающую значения t с вероятностью $\frac{1}{t}$ и 0 с вероятностью $\frac{t-1}{t}$.

Напомним, что достаточность условия была доказана в работе [4]. Теорема доказана.

Следствие 1. *Если в условиях теоремы для любых m и t справедливы неравенства $p_m(t) \geqslant \delta > 0$ для некоторого δ, то сходимость $\tilde{p}_m(t)$ по вероятности эквивалентна условию $\lim\limits_{t\to\infty} \sum\limits_{\tau=1}^{t} \omega_\tau^2(t) = 0$.*

Следствие 2. *При усреднении с экспоненциально убывающими весами и использовании «скользящего окна» последовательность $\tilde{p}_m(t)$ не сходится по вероятности.*

Автор выражает благодарность своему научному руководителю А. В. Галатенко за постановку задачи и внимание к работе.

Список литературы

[1] Александров Д. Е. Эффективные методы реализации проверки содержания сетевых пакетов регулярными выражениями // Интеллектуальные системы. — 2014. — Т. 18, вып. 1. — С. 37–60.

[2] Галатенко А. В. Активный аудит // JetInfo. — 1999. — № 8.

[3] Javitz H. S., Porras A. The NIDES statistical component description and justification. Technical report. — Computer Science Laboratory, SRI International, 1994.

[4] Галатенко А. В., Емельянов И. Н., Лебедев А. Е. Об обосновании алгоритмов статистического анализа в системах активного аудита // Материалы Четвертой международной кононференции «Проблемы безопасности и противодействия терроризму». — 2009. — Т. 2. — С. 195–209.