УДК 004.942:004.056

ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА ИНЦИДЕНТОВ В СИСТЕМЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ПРЕДПРИЯТИЯ ПРИ ПОМОЩИ ДИНАМИЧЕСКОЙ МОДЕЛИ

Заркумова-Райхель Р.Н., Абденов А.Ж.

ГОУ ВПО «Новосибирский государственный технический университет», Новосибирск, e-mail: zarkumova@gmail.com

Для прогнозирования количества инцидентов в системе информационной безопасности (ИБ) предприятия предлагается использовать линейную дискретную стохастическую модель в форме пространства состояний. Описывается алгоритм построения данной модели на основе использования экспертных оценок количества инцидентов ИБ. Для построения модели оцениваются дисперсия шумов динамики, дисперсия шумов наблюдений, дисперсия шума начального состояния исследуемого объекта. Постоянные коэффициенты уравнения модели рассчитываются на основе метода наименьших квадратов с учетом уровня зрелости предприятия с точки зрения обеспечения ИБ и при использовании алгоритма регуляризирующего кубического сплайна. Полученная модель позволяет, используя уравнения фильтра Калмана, рассчитать оценки предсказания и оценки фильтрации количества инцидентов в системе ИБ. Оценки фильтрации вычисляются путем корректировки оценок предсказания относительно данных наблюдений на момент времени предсказания.

Ключевые слова: инцидент информационной безопасности, уровень «зрелости» предприятия, модель в форме пространства состояний, фильтр Калмана

FORECASTING OF A NUMBER OF INFORMATION SECURITY INCIDENTS AT THE ENTERPRISE ON THE BASIS OF DYNAMIC MODEL

Zarkumova-Raykhel R.N., Abdenov A.Z.

NovosibirskStateTechnicalUniversity, Novosibirsk, e-mail: zarkumova@gmail.com

Linear discrete stochastic model in the form of state space is offered to use for forecasting of a number of information security incidents at the enterprise. The algorithm of this model formulation based on using expert judgements of a number of information security incidents is described. The variance of noise dynamics, the variance of noise observations, the variance of noise initial state of under study object are estimated for model formulation. The constant coefficients of model equation are calculated according to least-squares procedure and using an algorithm of regularizing cubic spline. Also the level of enterprise's maturity in terms of information security is taken into account when the constant coefficients are calculated. Based on resulting model, forecast estimations and filtration estimations of a number of information security incidents at the enterprise can be calculated by the Kalman's filter equations. Filtration estimations are calculated by adjusting the forecast estimations relative to the observation data on the instant forecast.

Keywords: information security incident, level of enterprise's maturity, model in the form of state space, the Kalman filter

В данной работе инцидентом ИБ будем считать любое незаконное, неразрешенное, неблагоприятное событие (НС), которое совершается в информационной системе.

Статистика инцидентов ИБ помогает осознавать количество инцидентов и их изменение во времени, определять наиболее актуальные угрозы для предприятия, прогнозировать количество инцидентов, и, тем самым, максимально точно планировать мероприятия по повышению уровня защищенности информационной системы предприятия.

В работе описывается подход к прогнозированию количества инцидентов на основании имеющейся статистики об инцидентах ИБ. Суть данного подхода заключается в том, что статистика инцидентов ИБ позволяет построить математическую модель системы ИБ, на основании которой вычисляются оценки предсказания количества НС в системе ИБ.

Постановка задачи

Запишем линейную стохастическую модель в форме пространства состояний [4]:

$$x(t+1) = f_i \cdot x(t) + g_i \cdot u(t) + w(t);$$

$$x_i(0) = x_{0i}\,; \quad i = 0,1,2,3; \qquad (1)$$

$$y(t+1) = x(t+1) + v(t+1), \; t = 0, 1, \ldots, \qquad (2)$$

где $x(t)$ – переменная, характеризующая количество инцидентов, зафиксированных в системе ИБ в момент времени t; $y(t)$ – наблюдаемое значение как измерение состояния исследуемого объекта, т.е. оценка количества НС в момент времени t; $\{u(t) = \text{const} = 4 - i\}$ – показатели, характеризующие пороговый уровень количества инцидентов ИБ для соответствующих значений i, где $i = \{0, 1, 2, 3\}$ – уровень «зрелости» предприятия с точки зрения обеспечения ИБ (далее – уровень «зрелости») [3];

$\{w(t),\ t = 0,\ 1,\ ...\}$ – гауссовская белая последовательность шумов динамики исследуемого объекта с нулевым средним и постоянной дисперсией Q_i; $\{v(t+1),\ t = 0,\ 1,\ ...\}$ – гауссовская белая последовательность шумов наблюдений с нулевым средним и постоянной дисперсией R_i; $x_i(0)$ – гауссовская случайная величина шума начального состояния исследуемого объекта с математическим ожиданием x_{0i} и дисперсией $P_i(0)$; f_i и g_i – коэффициенты в уравнении (1). Предполагается, что обе белые гауссовские последовательности $\{w(t), v(t+1), t = 0, 1, ...\}$ и $\{x_i(0), i = 0, 1, 2, 3\}$ взаимно независимы в пределах i-го уровня «зрелости».

Задача состоит в том, чтобы построить модель в форме пространства состояний, которая позволит с помощью уравнений фильтра Калмана осуществлять расчеты оценок предсказания, а также вычислять оценки фильтрации на основании корректировки оценок предсказания с учетом данных наблюдений на момент времени предсказания.

Решение задачи

Пусть известна статистика о количестве инцидентов ИБ, связанных с совершением пользователями ошибок при прохождении процедуры аутентификации.

Для построения математической модели в форме пространства состояний (1), (2), в первую очередь, необходимо определиться с уровнем «зрелости», воспользовавшись классификацией, предложенной компанией Gartner[1], которая подразумевает 4 уровня «зрелости» [3]. Самый низкий (нулевой) уровень ($i = 0$; $u(t) = 4 - i = 4$) характеризуется отсутствием понимания важности проблем ИБ, отсутствием финансирования; ИБ реализуется штатными средствами операционных систем. На предприятиях первого уровня ($i = 1$; $u(t) = 4 - i = 3$) финансирование ведется в рамках общего бюджета на информационные технологии; ИБ реализуется средствами нулевого уровня плюс средства резервного копирования, антивирусные средства, межсетевые экраны. На предприятиях второго уровня ($i = 2$; $u(t) = 4 - i = 2$) есть утвержденная руководством программа развития системы ИБ предприятия; финансирование ведется в рамках отдельного бюджета; ИБ реализуется средствами первого уровня плюс средства усиленной аутентификации, IDS, средства анализа защищенности, SSO, PKI и организационные меры. Самый высокий (третий) уровень ($i = 3$; $u(t) = 4 - i = 1$) предполагает, что ИБ является частью корпоративной культуры; финансирование ведется в рамках отдельного бюджета; ИБ реализуется средствами второго уровня плюс системы управления ИБ, CSIRT, SLA.

Таким образом, специалисты по ИБ или экспертная группа на конкретном предприятии могут зафиксировать значение входного управляющего сигнала $u(t)$, отнеся свою компанию к тому или иному уровню «зрелости» в соответствии с представленной классификацией.

Для построения модели вида (1), (2) необходимо вычислить коэффициенты f_i и g_i, дисперсию шумов динамики Q_i, дисперсию шумов наблюдений R_i и дисперсию шума начального состояния исследуемого объекта $P_i(0)$ для соответствующего уровня «зрелости» на основе имеющихся статистических данных.

Для того чтобы вычислить оценки дисперсий шумов динамики и шумов наблюдений для модели вида (1), (2), представим эту модель в упрощенной форме, которая позволит осуществить расчеты лишь на основе данных наблюдений объема выборки N:

$$x(t+1) = x(t) + w(t), \quad x(0) = x_0, \qquad (3)$$

$$y(t+1) = x(t+1) + v(t+1), \quad t = \overline{0, N-1}, \qquad (4)$$

где $x(t)$ – истинное значение состояния исследуемого объекта, $w(t)$ – шум динамики – случайная последовательность с нулевым средним и неизвестной дисперсией Q; $v(t+1)$ – шум наблюдений – случайная последовательность с неизвестными средним значением $E[v(t)] = q$ (систематическая ошибка) и неизвестной дисперсией R.

Для оценивания среднего значения q сформируем последовательности псевдоизмерений на основе двух и трех смежных данных наблюдений следующим образом:

$$v(t)^{(2)} = y(t) - y(t-1), \quad t = 2, 3, \ldots, N; \qquad (5)$$

$$v(t)^{(3)} = y(t) - \frac{1}{2}y(t-1) - \frac{1}{2}y(t-2), \quad t = 3, 4, \ldots, N. \qquad (6)$$

[1] Gartner – исследовательская и консалтинговая компания, специализирующаяся на рынках информационных технологий. Официальный сайт: www.gartner.com.

Оценка значения $\hat{q}$ в предположении о ее постоянстве определяется выражением:

$$\hat{q}(t\,|\,t) = \hat{q}(t-1\,|\,t-1) + \frac{1}{4(t+3)} abs(v(t)^{(2)} - \hat{q}(t-1\,|\,t-1)), \quad t = \overline{2, N}, \tag{7}$$

с начальным условием $\hat{q}(1|1) = 0$.

Среднее значение невязок:

$$E[v(t)^{(2)}] = q, \quad E[v(t)^{(3)}] = \frac{3}{2}q.$$

В [5] показано, что

$$E[(v(t)^{(3)} - \frac{3}{2}q)(v(t)^{(2)} - q)] = \frac{1}{2}Q.$$

Тогда последовательность псевдоизмерений дисперсии шумов динамики Q определяется следующим образом:

$$y(t)^{(w)} = 2(v(t)^{(3)} - \frac{3}{2}\hat{q}(t\,|\,t))(v(t)^{(2)} - \hat{q}(t\,|\,t)), \quad t = 3, 4, ..., N, \tag{8}$$

а оценку дисперсии $\hat{Q}$ можно рассчитать по рекуррентной формуле:

$$\hat{Q}(t\,|\,t) = \hat{Q}(t-1\,|\,t-1) + \frac{1}{N}(y(t)^{(w)} - \hat{Q}(t-1\,|\,t-1)), \quad t = \overline{3, N}, \tag{9}$$

с начальным условием $\hat{Q}(2\,|\,2) = 0$. Тогда при $t = N$ получим $\hat{Q}_i = \hat{Q}(N\,|\,N)$.

Также в [5] показано, что

$$E[(v(t)^{(2)} - q)^2] = 2R + Q.$$

Поэтому последовательность псевдоизмерений дисперсии шумов наблюдений R определяется следующим образом:

$$y(t)^{(v)} = \frac{1}{2}[(v(t)^{(2)} - \hat{q}(t\,|\,t))^2 - \hat{Q}(t\,|\,t)], \quad t = 2, 3, ..., N. \tag{10}$$

Таким образом, оценку дисперсии $\hat{R}$ можно рассчитать по рекуррентной формуле:

$$\hat{R}(t\,|\,t) = \hat{R}(t-1\,|\,t-1) + \frac{1}{t+1} abs(y(t)^{(v)} - \hat{R}(t-1\,|\,t-1)), \quad t = \overline{2, N}, \tag{11}$$

с начальным условием $\hat{R}(1|1) = 0$. Тогда при $t = N$ получим $\hat{R}_i = \hat{R}(N\,|\,N)$.

На основе метода наименьших квадратов [5] можно рассчитать коэффициенты $\hat{f}_i$ и $\hat{g}_i$ уравнения (1) в предположении, что $\tilde{x}(t) \approx y(t)$, $t = \overline{1, N}$, где $\tilde{x}(t)$ можно получить с помощью алгоритма регуляризирующего кубического сплайна [1].

Для окончательного построения линейной дискретной стохастической модели в форме пространства состояний вычисляется дисперсия шума начального состояния исследуемого объекта [2]:

$$P_0^{(t+1)} = \hat{f}_i \cdot P_0^{(t)} \cdot \hat{f}_i + \hat{Q}_i; \quad P_0^{(0)} = 0,$$

$$k = \overline{1, NN}, \quad \hat{P}_i(0) = P_i^{(NN)}, \tag{12}$$

где NN – число итераций, которое требуется для того, чтобы значение P_0 соответствовало значению оценки дисперсии установившегося начального состояния.

Полученные оценки дисперсий шумов $\hat{Q}_i$, $\hat{R}_i$ и $\hat{P}_i(0)$, значения начального состояния исследуемого объекта $x_i(0)$ и входного управляющего сигнала $u(t)$, коэффициенты $\hat{f}_i$ и $\hat{g}_i$ будут использоваться в математической модели, которая позволит рассчитать оценки предсказания и оценки фильтрации с помощью пяти уравнений фильтра Калмана (13)–(17) [4] при $i = \{0, 1, 2, 3\}$:

1) рассчитывается оценка предсказания $\hat{x}(t+1\,|\,t)$ количества инцидентов ИБ:

$$\hat{x}(t+1\,|\,t) = \hat{f}_i \cdot \hat{x}(t\,|\,t) + \hat{g}_i \cdot u(t), t = 0, 1, 2, ...;$$

$$x_i(0\,|\,0) = x_{0i}; \quad i = 0, 1, 2, 3; \tag{13}$$

2) рассчитывается дисперсия оценки предсказания состояния $P(t+1\,|\,t)$:

$$P(t+1\,|\,t) = \hat{f}_i^2 \cdot P(t\,|\,t) + \hat{Q}_i,$$

$$P_i(0|0); = P_{0i} \tag{14}$$

3) вычисляется коэффициент передачи фильтра $K(t+1)$:

$$K(t+1) = \frac{\hat{f}_i^2 \cdot P(t\,|\,t) + \hat{Q}_i}{\hat{f}_i^2 \cdot P(t\,|\,t) + \hat{Q}_i + \hat{R}_i} = \frac{P(t+1\,|\,t)}{P(t+1\,|\,t) + \hat{R}_i}; \tag{15}$$

4) рассчитывается оценка фильтрации $\hat{x}(t+1|t+1)$ количества инцидентов ИБ:

$$\hat{x}(t+1|t+1) = \hat{x}(t+1|t) + \\ + K(t+1)\cdot[y(t+1) - \hat{x}(t+1|t)]; \quad (16)$$

5) рассчитывается дисперсия оценки фильтрации $P(t+1|t+1)$:

$$P(t+1|\ t+1) = [1 - K(t+1)]\cdot P(t+1|\ t); \quad (17)$$

6) $t = t + 1$ и переход на шаг 1 и далее, пока $t \le N$.

Значения последовательностей шумов, оценки дисперсий шумов $\hat{Q}_i$, $\hat{R}_i$ и $\hat{P}_i(0)$, коэффициенты $\hat{f}_i$ и $\hat{g}_i$ могут принимать нецелые значения, поэтому оценки предсказания и фильтрации будут нецелочисленными. В связи с этим полученную оценку фильтрации необходимо округлить до ближайшего целого.

Рассмотрим пример.

Пример

Пусть условное предприятие с точки зрения «зрелости» информационной безопасности относится ко второму уровню «зрелости» ($i = 2$, $u(t) = 2$), и имеются экспертные оценки количества инцидентов за 2010 год (рис. 1).

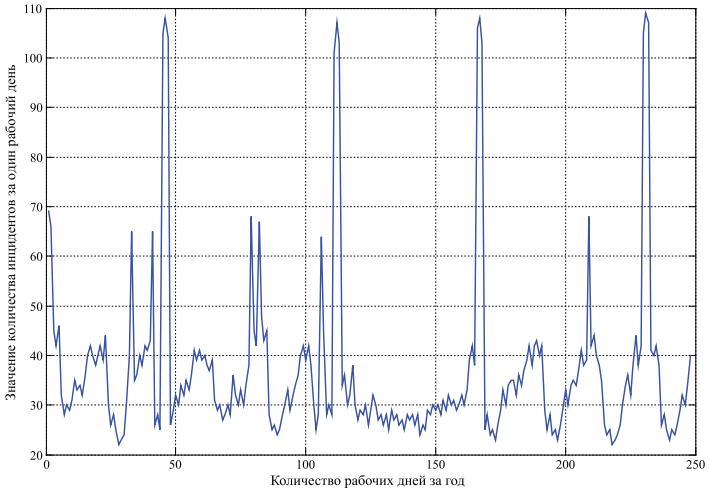

Рис. 1. Количество инцидентов ИБ, связанных с ошибками при входе в компьютерную систему, за весь 2010 год, за исключением нерабочих дней

По графику можно сделать вывод, что имеются четыре критические зоны с аномальными значениями количества инцидентов ИБ в диапазоне от 100 до 110 (связано с плановой сменой паролей) и семь критических зон с аномальными значениями в диапазоне от 60 до 70 инцидентов (по причине длительных выходных). Колебания количества НС в пределах от 20 до 50 инцидентов будем относить к «нормальному» режиму функционирования системы ИБ.

Учитывая «нормальное» поведение количества инцидентов и наличие критических зон, предлагается сгруппировать значения количества инцидентов следующим образом: [21–30], [31–40], [41–50], [61–70] и [101–110] инцидентов. Оценки характеристик модели необходимо вычислять для каждой вышеобозначенной группы.

Проведя расчеты согласно формулам (5)–(12), получим следующие оценки (табл. 1).

Таблица 1

Результаты расчета оценок f_2, g_2, Q_2, R_2 и $P_2(0)$относительно выделенных групп значений количества инцидентов

Коэффициенты	Значения коэффициентов в соответствующих группах				
	[21–30] инцидентов	[31–40] инцидентов	[41–50] инцидентов	[61–70] инцидентов	[101–110] инцидентов
$\hat{f}_2$	0,4373	0,4029	–0,1144	–0,4462	–0,1474
$\hat{g}_2$	7,6530	10,7120	23,8450	47,7692	60,4199
$\hat{Q}_2$	6,3528	11,0260	5,2286	5,9746	11,6764
$\hat{R}_2$	2,7773	5,4312	3,9343	1,5027	5,8256
$\hat{P}_2(0)$	7,8551	13,1628	5,2980	7,4594	11,9359

Теперь вычислим оценки предсказания и оценки фильтрации на основании уравнений фильтра Калмана (13)-(17), используя данные наблюдений за февраль 2011 г. (табл. 2). При этом будем учитывать наличие критических зон с аномальными значениями, т.е. использовать оценки характеристик модели, характеризующие уровень количества инцидентов в соответствующей группе значений.

На рис. 2 представлены результаты расчета оценок предсказания и фильтрации.

Таблица 2

Количество инцидентов ИБ в феврале 2011 г.

Рабочий день	1	2	3	4	5	6	7	8	9	10
Количество инцидентов	42	45	43	41	42	44	38	35	26	28
Рабочий день	11	12	13	14	15	16	17	18	19	
Количество инцидентов	25	22	23	24	30	29	70	35	37	

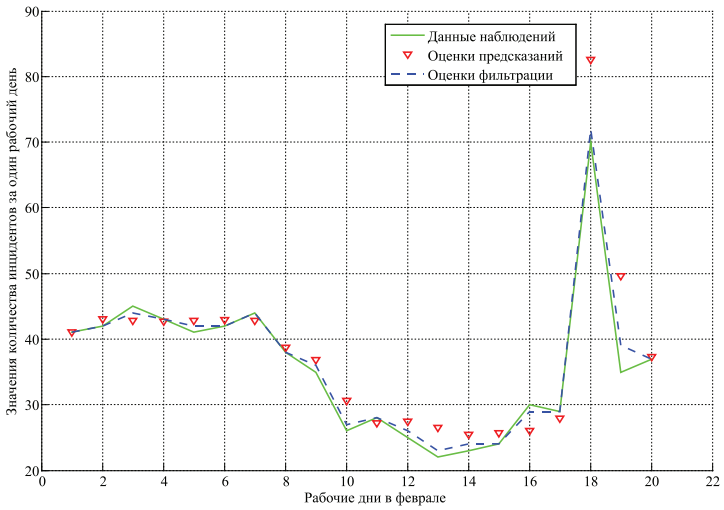

Рис. 2. Данные наблюдений, оценки предсказания и оценки фильтрации количества инцидентов в феврале 2011 г.

Из рисунка видно, что рабочие дни в феврале относительно числа инцидентов визуально можно разделить на 5 подынтервалов: 1 подынтервал со значениями количества НС, входящими в диапазон с 41 по 50 инцидентов; 2 – с 31 по 40; 3 – с 21 по 30; 4 – с 61 по 70; 5 – с 31 по 40 инцидентов. Для расчета оценок предсказания и фильтрации в каждом подынтервале использовались наборы оценок f_2, g_2, Q_2, R_2 и $P_2(0)$, характерные для соответствующих групп значений количества инцидентов. Согласно рис. 2 оценки фильтрации более приближены к значениям данных наблюдений, что подчеркивает относительную достоверность данных наблюдений по сравнению с оценками предсказания.

Заключение

В данной работе была предложена методика построения математической модели системы информационной безопасности в виде линейной дискретной стохастической модели в форме пространства состояний относительно количества инцидентов ИБ конкретного «уровня» зрелости. Значения количества инцидентов были сгруппированы, учитывая «нормальное» поведение количества инцидентов и наличие критических зон с аномальными значениями количества НС. Для каждой группы значений количества НС была построена соответствующая линейная дискретная стохастическая модель в форме пространства состояний. Все полученные модели позволили рассчитать оценки фильтрации на основании оценок предсказания и текущих наблюдений на момент времени предсказания. При этом оценки фильтрации более достоверно оценивают состояние объекта относительно количества инцидентов, происходящих в компьютерной системе для соответствующего уровня «зрелости» исследуемого объекта.

Список литературы

1. Абденов А.Ж., Абденова Г.А., Снисаренко А.В. Построение и применение кубических сплайнов для сглаживания и дифференцирования данных наблюдений: методическое пособие. – Новосибирск: Изд-во НГТУ, 2004. – 31 с.

2. Абденова Г.А. Прогнозирование значений уровня временного ряда на основе уравнений фильтра Калмана // Ползуновский вестник. – 2010. – №2. – С. 4–6.

3. Информационная безопасность. Уровни зрелости СОИБ организации [Электронный ресурс]. – Режим доступа: http://www.wikisec.ru (дата обращения: 17.03.2012).

4. Синицын И.Н. Фильтры Калмана и Пугачева // Рос. акад. науки, Ин-т проблем информатики. – М.: Логос, 2007. – 772 с.

5. Mehra R. Identification and adaptive Kalman filtering // Mechanics. – 1971. –м № 3. – P. 34–52.

References

1. Abdenov A.Zh, Abdenova G.A., Snisarenko A.V. *Postroenie i primenenie kubicheskikh splaynov dlya sglazhivaniya i differentsirovaniya dannykh nablyudeniy* (Construction and application of cubic splines for smoothing and differentiation of the observation data). Novosibirsk, NSTU Publishing House, 2004, 31 p.

2. Abdenova G.A. *Polzunovskiy vestnik*, 2010, no.2, pp. 4–6.

3. *Informatsionnaya bezopasnost'. Urovni zrelosti SOIB organizatsii* (Information Security. Levels of enterprise's maturity). Avalaible at: http://www.wikisec.ru (accessed 23 March 2012).

4. Sinitsyn I.N. *Fil'try Kalmana i Pugacheva* (Filters of Kalman and Pugachev). Moscow, Logos, 2007. 772 p.

5. Mehra R. *Mechanics*, 1971, no. 3, pp. 34–52.

Рецензенты:

Фионов А.Н., д.т.н., проректор по научной работе Сибирского государственного университета телекоммуникаций и информатики, г. Новосибирск;

Разинкин В.П., д.т.н., профессор кафедры теоретических основ радиотехники Новосибирского государственного технического университета, г. Новосибирск.

Работа поступила в редакцию 18.05.2012.