УДК 004.056(06)

Н.А. Богульская, М.М. Кучеров, Е.А. Кресан

Защита персональных данных в информационных системах

Рассматриваются вопрос внедрения смарт-карт для идентификации личности в образовательном учреждении и связанные с этим вопросы защиты персональных данных.
Ключевые слова: идентификация, базы персональных данных, смарт-карты.

Управление информационными ресурсами означает, что руководство предприятия признает, что информация является важнейшим ресурсом и предпринимает соответствующие действия по управлению ею. Они включают в себя:

– определение информационной базы через элементы данных, существенные для производственного процесса (важная информация);

– оценку или классификацию данных для того, чтобы позволить предпринять соответствующие меры защиты.

Персональные данные можно рассматривать как особое подмножество конфиденциальной информации [1]. В формальных моделях безопасности классификация уровней безопасности состоит из уровня, отражающего значение конфиденциальности/целостности, и (возможно, пустого) множества подмножеств, обычно называемых категориями. Обычно уровни расположены в линейном порядке по конфиденциальности или целостности, образуя частично-упорядоченную решетку [2].

Предлагается вместо уровней также рассмотреть категории, а именно множества подмножеств атрибутов конфиденциальности и целостности, взяв за основу двойную решетку Белнапа с четырьмя истинностными значениями «T, F, Both (T, F), None». Смысл значений состоит в следующем: атрибут является только истинным (T); только ложным (F); одновременно истинным и ложным, например, по разным источникам или в разные моменты времени (Both, {T, F}); или имеющим неизвестный статус, т.е. ни истинным, ни ложным (None, N) [3].

Причина использования категорий заключается в том, что разные источники информации могут обеспечивать несогласованные данные, а их сочетание может привести к скрытым противоречиям.

Рассмотрим, например, универсальное идентификационное устройство для студентов, преподавателей и других сотрудников университета на основе бесконтактных смарт-карт и считывателей. Для защиты смарт-карт от копирования используются временные ключи. Для этого надо разделить ключ аутентификации на три части. Одна (постоянная) часть ключа хранится на сервере, вторая (постоянная) часть ключа является серийным номером карты, а третья (переменная) часть ключа хранится в памяти карты. Переменная часть ключа является меткой времени последней аутентификации, зашифрованной на ключе сервера. Считыватель принимает серийный номер карты и отправляет его на сервер, сервер проверяет права доступа, если объекту доступ разрешен, то он запрашивает переменную часть ключа, хранящегося в памяти карты. Получив данные, сервер расшифровывает их на своем ключе и проверяет в базе данных, совпадают ли метки времени. Если они совпадают, то доступ разрешен, иначе сервер заносит серийный номер карты в черный список. Далее сервер зашифровывает текущую метку времени и отправляет ее на считыватель, который перезаписывает переменную часть ключа.

Защита персональных данных

Внедрение смарт-карт представляет собой применение так называемых приватно-инвазивных технологий, широкое использование которых привлекло внимание к вопросу о неприкосновенности персональной информации. В работе [4] указано, что «конфиденциальность персональных данных представляет собой заинтересованность лица в управлении обработкой или, по меньшей мере, в серьезном влиянии на обработку данных о себе», таким образом, область неприкосновенности персональной информации выходит за рамки конфиденциальности, затрагивая этику.

Информацию в базе данных можно разделить на неидентифицируемую информацию, которая не имеет персонального характера (НИИ), и персональные данные (ПД). Последние могут включать персональную идентификационную информацию (фамилию, имя, отчество, серию и номер паспорта) или содержать неидентифицируемые данные (год, месяц, дату рождения, пол). Необходимо рас-

смотреть отличия, связанные с ПД, между различными типами информации для доказательства того, что безопасность, политика и технические требования, которые диктуют персональные данные, приводят к необходимости для базы данных ПД своей концептуальной схемы, системы управления, а также физической структуры. В частности, выделяются:

– ПИИ (персонально-идентифицируемая информация) – информация, по которой можно установить личность; и

– ПНИ (персонально-неидентифицируемая информация) – множество элементов, не позволяющих самостоятельно идентифицировать личность. ПНИ – подмножество неидентифицируемой информации (НИИ), которая, однако, называется персональной, поскольку субъект ПД заинтересован в ее конфиденциальности. Эта информация составляет информационную область субъекта ПД.

В отличие от НИИ для ПИИ используются особая технология и методы (например, обезличивание). Кроме того, Федеральный закон РФ от 27 июля 2006 г. №152-ФЗ позволяет отделить ПД от других видов информации, что облегчает их организацию в порядке, не доступном для других видов информации.

Классификация элементов информации

Рассмотрим подход к классификации элементов информации (ЭИ) на основе модели ценности [5]. В ней каждый ЭИ рассматривается с точки зрения его субъективного или объективного значения. В первом случае речь идет о его конфиденциальности, во втором – о его целостности и доступности. Некоторые ЭИ одновременно имеют как субъективную, так и объективную ценность.

Любая двойная решетка представляется как структура, упорядоченная по двум базовым свойствам, которые являются существенными для элементов, формирующих решетку. Первым свойством является содержание информации. Природа ЭИ несущественна, более важным является содержание информации. Удаление из персональных данных ПИИ, а из НИИ, соответственно, ПНИ, уменьшает количество информации в ЭИ. Для информационной безопасности вторым свойством, очевидно, является конфиденциальность.

Существует ряд способов, с помощью которых можно получить вероятностные границы для подмножеств конфиденциальности и целостности и связать их с ЭИ (например, экспертная оценка, статистика). Закон сетевой экономики Гроша утверждает, что прибыль от использования компьютерной системы возрастает как корень квадратный из ее мощности, т.е. для того, чтобы сделать вычисления в 10 раз дешевле, надо выполнять их в 100 раз быстрее. В соответствии с этим ЭИ, имеющие высокую доступность, должны обрабатываться на нижних уровнях иерархии конфиденциальности. Следовательно, НИИ должна находиться в основании иерархии по конфиденциальности и обладать высокой степенью надежности.

Подход, объединяющий рассмотрение ПД и НИИ, делает возможной работу с ЭИ, которые можно характеризовать не только с точки зрения конфиденциальности, целостности или эффективности доступа, но также по наполнению информационным содержанием и как объекты, обладающие субъективной/объективной ценностью.

Следует остановиться на том, что указывать в качестве параметров: точные оценки или интервалы? Эксперты чаще указывают на интервалы. Кроме того, известно [6], что если имеется недостаточно информации относительно взаимодействия между двумя событиями, то вероятность совместного события не может быть определена точно: можно дать только строгие границы. Следовательно, ЭИ можно связать с интервалами вероятностей. Это приводит к следующему определению для категорий:

Определение 1 (доверительный интервал, ДИ)

Обозначим $E[0, 1]$ множество всех замкнутых подынтервалов на [0, 1]. Рассмотрим множество $L_c = E[0,1]\times E[0,1]$. Доверительный интервал – элемент множества L_c. Обозначим его как $\langle[\alpha, \beta], [\gamma, \delta]\rangle$.

Интервалы связаны с «уровнями» ЭИ. Значение интервала $CL(A) = \langle[\alpha, \beta], [\gamma, \delta]\rangle$ для сущности A состоит в том, что α и β представляют нижнюю и верхнюю вероятностную границу заданной оценки ПИИ в A, а γ и δ – нижняя и верхняя вероятностная граница оценки для НИИ. Обычно $\alpha \le \beta$, а $\gamma \le \delta$.

Определение 2 (*порядок*)

Рассмотрим множество $L_c = E[0,1] \times E[0,1]$. Пусть $\langle[\alpha_1,\beta_1], [\gamma_1,\delta_1]\rangle$ и $\langle[\alpha_2,\beta_2], [\gamma_2,\delta_2]\rangle$ – произвольные элементы множества L_c, тогда

$\langle[\alpha_1, \beta_1],[\gamma_1, \delta_1]\rangle \le_s \langle[\alpha_2, \beta_2],[\gamma_2, \delta_2]\rangle$, если $\alpha_1 \le \alpha_2$, $\beta_1 \le \beta_2$ и $\gamma_2 \le \gamma_1$, $\delta_2 \le \delta_1$,

$\langle[\alpha_1, \beta_1],[\gamma_1, \delta_1]\rangle \leq_k \langle[\alpha_2, \beta_2],[\gamma_2, \delta_2]\rangle$, если $\alpha_1 \leq \alpha_2$, $\beta_2 \leq \beta_1$ и $\gamma_1 \leq \gamma_2$, $\delta_2 \leq \delta_1$,

$\langle[\alpha_1, \beta_1],[\gamma_1, \delta_1]\rangle \leq_f \langle[\alpha_2, \beta_2],[\gamma_2, \delta_2]\rangle$, если $\alpha_1 \leq \alpha_2$, $\beta_2 \leq \beta_1$ и $\gamma_2 \leq \gamma_1$, $\delta_1 \leq \delta_2$.

Верхние и нижние элементы по отношению к разным порядкам следующие. Индекс указывает на соответствующий порядок.

$T_s = \langle[1,1], [0,0]\rangle$ или Tt, $\quad \perp_s = \langle[0,0], [1,1]\rangle$ или Ff,

$T_k = \langle[1,0], [1,0]\rangle$ или TFtf, $\quad \perp_k = \langle[0,1], [0,1]\rangle$ или N,

$T_f = \langle[1,0], [0,1]\rangle$ или TF, $\quad \perp_f = \langle[0,1], [1,0]\rangle$ или tf.

Символ T_s соответствует оценке «ПИИ, отсутствует НИИ», а $\perp_s$ – «НИИ, отсутствует ПИИ». Символ T_k соответствует противоречивой оценке информации, которая делает интервалы пустыми (в операционном смысле), а $\perp_k$ соответствует наименее определенной информации [0,1], так как задает только тривиальные границы по вероятностям. Символ T_f соответствует оценке только субъективной ценности ЭИ, а символ $\perp_f$ – только его объективной ценности. Первые два порядка сопоставимы с порядками в двойных решетках. Последний порядок не имеет аналогий. Интуитивно информация приобретает большую субъективную ценность в связи с наличием в ней ПД.

Определение 3 (*алгебра уровней*)

Пусть $\left\langle L_C, \leq_s, \leq_k, \leq_f \right\rangle$ дано согласно определению 1. Имеем следующие операции пересечения и объединения, соответствующие частичным порядкам:

1. $\left\langle[\alpha_1,\beta_1],[\gamma_1,\delta_1]\right\rangle \otimes_s \left\langle[\alpha_2,\beta_2],[\gamma_2,\delta_2]\right\rangle = \langle\left[min\{\alpha_1,\alpha_2\}, min\{\beta_1,\beta_2\}\right], \left[max\{\gamma_1,\gamma_2\}, max\{\delta_1,\delta_2\}\right]\rangle.$
2. $\left\langle[\alpha_1,\beta_1],[\gamma_1,\delta_1]\right\rangle \oplus_s \left\langle[\alpha_2,\beta_2],[\gamma_2,\delta_2]\right\rangle = \langle\left[max\{\alpha_1,\alpha_2\}, max\{\beta_1,\beta_2\}\right], \left[min\{\gamma_1,\gamma_2\}, min\{\delta_1,\delta_2\}\right]\rangle.$
3. $\left\langle[\alpha_1,\beta_1],[\gamma_1,\delta_1]\right\rangle \otimes_k \left\langle[\alpha_2,\beta_2],[\gamma_2,\delta_2]\right\rangle = \left\langle\left[min\{\alpha_1,\alpha_2\}, max\{\beta_1,\beta_2\}\right], \left[min\{\gamma_1,\gamma_2\}, max\{\delta_1,\delta_2\}\right]\right\rangle.$
4. $\left\langle[\alpha_1,\beta_1],[\gamma_1,\delta_1]\right\rangle \oplus_k \left\langle[\alpha_2,\beta_2],[\gamma_2,\delta_2]\right\rangle = \left\langle\left[max\{\alpha_1,\alpha_2\}, min\{\beta_1,\beta_2\}\right], \left[max\{\gamma_1,\gamma_2\}, min\{\delta_1,\delta_2\}\right]\right\rangle.$
5. $\left\langle[\alpha_1,\beta_1],[\gamma_1,\delta_1]\right\rangle \otimes_f \left\langle[\alpha_2,\beta_2],[\gamma_2,\delta_2]\right\rangle = \left\langle\left[min\{\alpha_1,\alpha_2\}, max\{\beta_1,\beta_2\}\right], \left[max\{\gamma_1,\gamma_2\}, min\{\delta_1,\delta_2\}\right]\right\rangle.$
6. $\left\langle[\alpha_1,\beta_1],[\gamma_1,\delta_1]\right\rangle \oplus_f \left\langle[\alpha_2,\beta_2],[\gamma_2,\delta_2]\right\rangle = \left\langle\left[max\{\alpha_1,\alpha_2\}, min\{\beta_1,\beta_2\}\right], \left[min\{\gamma_1,\gamma_2\}, max\{\delta_1,\delta_2\}\right]\right\rangle.$

Алгебраические свойства интервалов и решеток, на которые они опираются, имеют самостоятельный интерес и могут использоваться для развития моделей доступа.

Интерпретация модели

Возможны три разных измерения для ограничения доступа к информации и информационным процессам на основе конфиденциальности, количества информации и субъективной/объективной ценности информации соответственно. Рассмотрим определение 3. Правило 1 описывает изменения в объектах и касается информации, вставляемой, например, в объект 2. Она должна иметь классификацию ниже классификации этого объекта. Анализ показывает эквивалентность правила 1 правилу модели контроля конфиденциальности: не позволяется размещать информацию или записывать ее в объекты, имеющие более низкий уровень секретности. Правило 2 соответствует чтению из объекта 1. Правило 2 можно интерпретировать как запрет на просмотр объектов, классификация которых превышает «уровень доверия» субъекта 2.

Если новый документ создан путем удаления из него чувствительной информации, ПИИ, а также удаления всей ПНИ из НИИ, то новый документ имеет неполную информацию и может быть отнесен к более низкому уровню. Правило 3 не допускает утечку информации, например дополнение низкоуровневого документа за счет ПИИ. Правило 4 связано с чтением документов.

Правила 5 и 6 позволяют рассматривать ПИИ и НИИ независимо друг от друга. Правила классификации на основе теории вероятностей более подробно обсуждаются в [7].

Полученный аппарат можно применить к базам ПД, которые содержат персональную информацию и информацию, с ней связанную. Целью атак на ПИИ, таких как data-mining, является установление личности субъекта при помощи неидентифицируемой информации, например определение идентичности из обезличенной информации, касающейся возраста, пола и почтового индекса. Выделение ПИИ из НИИ позволяет упростить сложную политику, необходимую для защиты конфиденциальной информации, там, где применяется несколько правил.

Центральный механизм БД ПД представляет собой объявление субъектов в таблице, называемой СУБЪЕКТЫ ПД (СПД), которая включает в себя уникальные идентификаторы всех субъектов в БД ПД. Таблица СПД содержит отдельную запись с внутренним ключом (# субъекта) для каждого

субъекта ПД в дополнение к другой информации, такой как указатель на его информационную область ПД (ОПД).

Принцип уникальности идентификаторов субъектов требует, чтобы внутренний ключ (# субъекта) отображался по принципу один-к-одному на идентичность лица или физическое местонахождение объекта. Для БД ПД, в которой имеются три записи, каждая ведет к трем ОПД:

PROPRIETOR_TABLE:

((# proprietor1, CL_1) $\Rightarrow$ ОПД субъекта 1,

(# proprietor2, CL_2) $\Rightarrow$ ОПД субъекта 2,

(# proprietor3, CL_3) $\Rightarrow$ ОПД субъекта 3),

где CL_i означает доверительный интервал. Схема позволяет провести полную изоляцию ПИИ. В приведенном выше примере можно назначить независимую политику доступа для трех ОПД.

Базы данных ПД, в которые включены ДИ, подчиняются всем предложениям, определенным для ДИ. В качестве иллюстрации применения этих положений рассмотрим случай ограничения конфиденциальности ПД, запрещающий раскрывать ПИИ. Смешивание (например, изменение, вставка и т.д.) ПИИ с любым другим фрагментом информации делает ограничение раскрытия распространяющимся на комбинированную информацию. В этом случае общая политика состоит в том, что применение правила защиты к $\sigma_1 \in$ ПИИ предполагает применение такой же защиты к $(\sigma_1 \otimes \sigma_2)$, где $\sigma_2 \notin$ ПИИ.

Пример: монитор безопасности пересылок

Модель контроля информационных потоков – это модель, в которой монитор безопасности пересылок (МБП) принимает решение о доступе субъекта к объекту по соответствующему запросу субъекта на доступ или запрещает его. Предположим, что система содержит информацию из двух классов: ПИИ и НИИ и что МБП имеет два блока; $МБП_s$, который контролирует доступ к секретной информации, и $МБП_o$, который подключает пользователя к банку с НИИ (рис. 1).

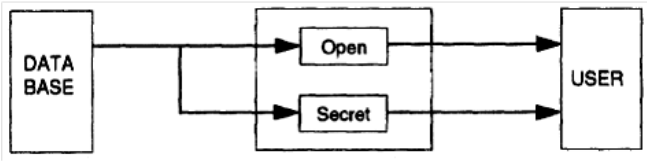

Рис. 1. Монитор безопасности пересылок для управления доступом

Блок $МБП_s$ имеет два вида отказов: режим 1, в котором прекращается доступ к ПИИ, и режим 4, в котором разрешается доступ к персональным данным. Блок $МБП_o$ имеет только один вид отказов: режим 2, в котором прекращается доступ к НИИ. Допустим, что «старение» системы можно моделировать как случайный марковский процесс с дискретными состояниями [8]. При анализе поведения системы во времени удобно пользоваться графом состояний, содержащим столько вершин, сколько различных состояний возможно у системы.

Ребра графа состояний отражают возможные переходы из некоторого состояния во все остальные в соответствии с параметрами потоков отказов. Если для каждого состояния изделия, другими словами, для каждой вершины графа вычислить вероятность нахождения изделия именно в этом состоянии в любой произвольный момент времени $P_i(t)$, то, зная эти вероятности, можно оценить интересующие на практике показатели надежности на основе матрицы смежности.

Матрица смежности графа состояний МБП представляет собой

$$\begin{matrix} L_0 \\ L_1 \\ L_2 \\ L_3 \\ L_4 \end{matrix} \begin{pmatrix} 0 & W_2 & W_1 & 0 & W_3 \\ 0 & 0 & 0 & W_1 & 0 \\ 0 & 0 & 0 & W_2 & W_3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

где L_0 – начальное состояние, а L_i $(i \neq 0)$ – состояния, в которых наблюдаются отказы.

Состояние МБП можно формально записать следующим образом. Состояние L_0 можно обозначить как $CL_0 = \langle[1,0],[1,0]\rangle = T_k$. В этом случае согласно определению 3 возможна запись в него как ПИИ, так и ПНИ. Обозначим $CL_1 = \langle[0,1],[1,0]\rangle = \perp_f$. Здесь возможна только запись ПНИ. Далее, $CL_2 = \langle[1,0], [0,1]\rangle = T_f$. Здесь возможна только запись ПИИ. В состоянии $CL_3 = \langle[0,1], [0,1]\rangle = \perp_k$ прекращается пересылка любой информации в МБП. И, наконец, состояние L_4, в котором имеется утечка информации, может быть связано с отказом системы санкционирования доступа.

Непосредственные вычисления приводят к вероятности нахождения МБП в состоянии L_4:

$$P_4(t) = \frac{W_3}{W_2 + W_3}\left(1 - e^{-(W_2+W_3)t}\right).$$

Подстановка числовых выражений $W_1 = W_2 = 9{,}5/10000$ отказов/ч и $W_3 = 1/10000$ отказов/ч приводит к оценке $1/(W_2 + W_3) = 932$ (ч).

Заключение

Каждый элемент персональных данных связывается с определенным доверительным интервалом, показывающим вероятностные границы заданной оценки конфиденциальности и целостности, который может быть определен экспертом или базироваться на собранной статистике. Доверительные интервалы имеют в своей основе алгебраическую структуру, называемую тройной решеткой. В частности, тройная решетка позволяет упорядочить персональные данные, предлагая порядок по конфиденциальности, количеству и субъективной/объективной ценности информации.

Литература

1. Федеральный закон РФ от 25 июля 2011 г. № 261-ФЗ «О внесении изменений в Федеральный закон «О персональных данных» [Электронный ресурс]. – Режим доступа: http://www.rg.ru/gazeta/rg/2011/07/27.html, свободный (дата обращения: 30.04.2012).

2. Основы информационной безопасности: учеб. пособие / Е.Б. Белов, В.П. Лось, Р.В. Мещеряков, А.А. Шелупанов. – М.: Горячая линия – Телеком, 2006. – 544 с.

3. Belnap N.J. A useful four valued logic // Modern Uses of Multiple-Valued Logic, Dunn, J.M., Epstain, G. (eds.). – D. Reidel Publishing Company, Dordrecht, 1977. – P. 8–37.

4. Kang J. Information privacy in cyberspace transactions // Stanford Law Review. – 1998. – Vol. 1193. – P. 1212 – 1220.

5. Schweitzer J.A. Protecting Information on Local Area Networks. Boston: Butterworth-Heinemann, 1988. – P. 28–30.

6. Fenstad J. E. The structure of probabilities defined on first-order languages // Studies in Inductive Logic and Probabilities, Jeffrey, R.C. (ed). – University of California Press, 1980. – Vol. 2. – P. 251–262.

7. Защита персональных данных в информационных системах / Н.А. Богульская, М.М. Кучеров, А.А. Хохлов // IV Пленум СибРОУМО: (материалы Пленума); XII Всерос. науч.-практ. конф. «Проблемы информационной безопасности государства, общества и личности»: (докл. конф.) 8–13 июня 2010 г., Томск–Барнаул–Белокуриха: науч.-метод. и нормативные материалы и документы. – Томск: В-Спектр, 2010. – С. 188–195.

8. Панфилов И.В. Вычислительные системы / И.В. Панфилов, А.М. Половко. – М.: Сов. радио, 1980. – 304 с.

Богульская Нина Александровна
Канд. физ.-мат. наук, доцент каф. прикладной математики и компьютерной безопасности ИКИТ СФУ
Тел.: 8 (391) 243-92-21
Эл. почта: NBogulskaya@sfu-kras.ru

Кучеров Михаил Михайлович
Канд. физ.-мат. наук, доцент каф. прикладной математики и компьютерной безопасности ИКИТ СФУ
Тел.: 8 (391) 243-92-21
Эл. почта: MKucherov@sfu-kras.ru

Кресан Евгений Александрович
Аспирант каф. прикладной математики и компьютерной безопасности ИКИТ СФУ
Тел.: 8 (391) 243-92-21
Эл. почта: Kresan@mail.ru

Bogul'skaya N.A., Kucherov M.M., and Kresan Ye.A.

Security of personal data in information systems

It is considered the question of introduction of smart cards for identification at an educational institute and the related issues on the protection of personal data.

Keywords: identification, personal databases, smart cards.