ББК Х303.5

ОЦЕНКА АКТУАЛЬНОСТИ ЗАЩИЩАЕМОЙ ИНФОРМАЦИИ

С.Г. Фомичева

Норильский индустриальный институт

Рецензент В.Е. Подольский

Ключевые слова и фразы: достоверность информации; защита информации; интеллектуальная собственность; схема информационно-аналитической системы; ценность информации.

Аннотация: Предпринята попытка формализации актуальности информации и ее защиты. Решения проблем интеллектуальной собственности сейчас находятся в стадии становления. Очевидно, что защитить весь объем информации невозможно. Несомненно, что защита информации требует дополнительных как стоимостных, так и временных ресурсов.

Современное общество характеризуется колоссальным ростом информации, значимость которой выходит на первый план при принятии деловых решений и, по сути, определяет развитие экономики. На сегодняшний день начинают классифицировать страны по уровню их развития в соответствии со следующими категориями [1]:

– *развитые* – страны, способные производить и продавать информационные услуги;

– *развивающиеся* – страны, не производящие информационных услуг (на продажу), но создающие и продающие материальные товары (машины, бытовую технику, самолеты и др.);

– *неразвитые* – страны, не производящие ни того, ни другого, и являющиеся поставщиками сырья и рабочей силы для стран более высоких категорий. Они же являются покупателями соответствующей продукции ведущих стран.

Несомненно, информация стала фактически международным товаром со всеми свойственными любому товару качествами, такими как затраты на производство, прибыль, цена и т.д. Однако, продавая информацию, ее владелец сам не лишается этой информации, она остается в его собственности (т.е. информация обладает свойством «неотчуждаемости»), что, в свою очередь, кардинально меняет, в отношении информации, такие поня-

Фомичева С.Г. – кандидат технических наук, доцент, заведующая кафедрой «Информационные системы и технологии» Норильского индустриального института.

тия, как ее кража или порча. Решения проблем интеллектуальной собственности сейчас находятся в стадии становления. Очевидно, что в этом контексте важнейшими становятся задачи криптозащиты (от несанкционированного прочтения) и имитозащиты (от подмены) информации. Также очевидно, что защитить весь объем информации невозможно. Несомненно, что защита информации требует дополнительных как стоимостных, так и временных ресурсов. Как априорно определить, *какую* информацию необходимо защищать? С какой *степенью* защиты? Можно ли формализовать оценку *актуальности* информации и затем динамично *управлять* процессом защиты информации?

В данной статье автором предпринята попытка формализации актуальности информации. Для этого введем ряд определений.

Актуальность информации – *ценность* информации в *данный* момент времени. Ценность информации всегда субъективна, поскольку одна и та же информация для одного субъекта (организации, человека) особо важна, а для другого – бесполезна, хотя количество информации (определяемое через энтропию) для этих субъектов одинаковое. Однако это совершенно не означает, что энтропия и ценность информации являются независимыми параметрами. Как отмечает Г.П. Шанкин [1], количество информации и ее ценность находятся в сложном отношении. Он же дает определение ценности информации (модель ценности информации – α, β-модель).

Пусть информация вызывает определенное действие пользователя *User*. Эффективность этого действия поддается некоторой количественной оценке $\alpha > 0$. При отсутствии информации *User* предпринял бы некоторое другое действие, эффективность которого есть $\beta > \alpha$. Тогда *ценность* полученной информации есть величина, пропорциональная разности α–β. То есть *ценность* информации трактуется как *эффективность* действий, основанных на *этой* информации, при предположении, что *User* использует информацию наиболее эффективным образом.

Будем рассматривать схему информационно-аналитической системы, предложенную Г.П. Шанкиным (рис. 1). Имеется объект наблюдения *Object*, который может находиться в одном из состояний конечного множества *O*. В фиксированный момент времени *Object* находится в состоянии $o^* \in O$. Состояние объекта изучает наблюдатель *Tracer*. Эти наблюдения в общем случае могут быть неточными в следующем контексте: *Tracer* не наблюдает «истинные» состояния, а лишь выделяет наиболее вероятное подмножество $O' \subseteq O$, в котором находится $o^* \in O$. Результат своих наблюдений *Tracer* формулирует в некотором сообщении $x \in X$, где X – конечное множество возможных сообщений. Содержанием сообщения $x \in X$ является множество $O(x) \subseteq O$, которому, по мнению *Tracer,* принадлежит истинное состояние $o^* \in O$. Сообщение $x \in X$ *Tracer* направляет по некоторому каналу связи (*Channel*) пользователю информации *User*. *User* использует эту информацию для достижения некоторой цели, связанной с его знаниями о состоянии $o^* \in O$.

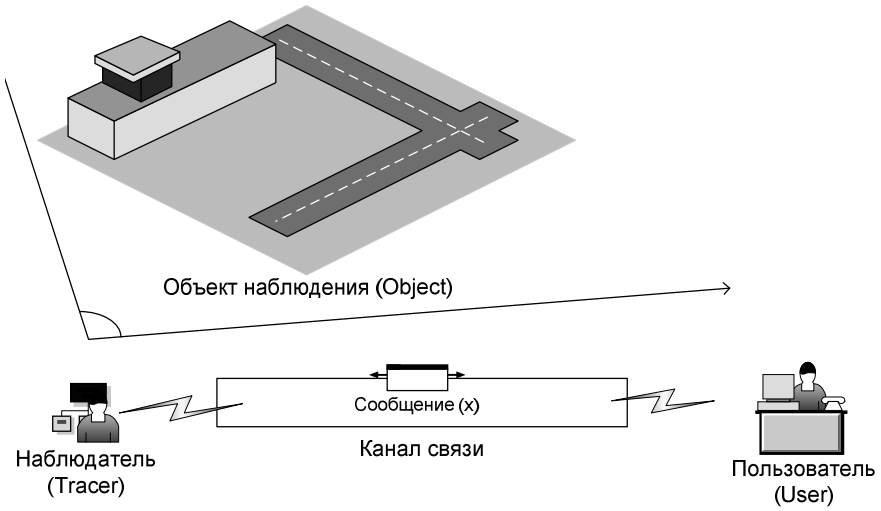

Рис. 1 Структурная схема информационно-аналитической системы

Действия *User* описываются следующим образом. Исходя из своих априорных знаний и полученного сообщения, *User* стремится выделить наиболее вероятное состояние $o' \in O$, в котором находится *Object*. Выделив это состояние, *User* предпринимает некоторые действия, эффективность которых может быть измерена количественно. Если *User* верно «угадал» состояние *Object* (т.е. $o' = o^*$), то эффективность оценивается величиной $\alpha > 0$, если же он не угадал ($o' \neq o^*$), то эта эффективность есть $\beta < \alpha$. Это справедливо для всех o^*, $o' \in O$, величины α, β не зависят от конкретных o^*, $o' \in O$ и являются константами.

В данной модели ценности информации предполагается, что искажения и задержки в *Channel* отсутствуют, а *Tracer* и *User* одинаково понимают содержание $O(x)$ сообщения $x \in X$ (имеет место полное взаимопонимание *Tracer* и *User*). *Tracer* формирует и посылает *User* достоверные сообщения, т.е. $o^* \in O$.

Обозначим

$$X^+ = \left\{ x \in X \,\middle|\, o^* \in O(x) \right\}.$$

$$X^- = X \setminus X^+.$$

Достоверность сообщения $x \in X$ означает, что $x \in X^+$. Все состояния $o' \in O$ равноправны как «претенденты» на истинное состояние $o^* \in O$, т.е. вероятность $P\left\{ o' = o^* \,\middle|\, o' \in O(x) \right\} = \left| O(x) \right|^{-1}$, $x \in X^+$. До получения сообщения $x \in X$ от *Tracer User* обладает априорной информацией о состоянии *Object*, которую можно представить в виде наличия у него априорного со-

общения $x_0 \in X$. Сообщение $x_0 \in X$ достоверно $o^* \in O(x_0)$, и все состояния $o' \in O(x_0)$ равноправны как претенденты на $o^* \in O$, т.е.

$$P\left\{o' = o^* \middle| o' \in O(x_0)\right\} = \left|O(x_0)\right|^{-1}. \quad (1)$$

Ценность информации определяется *User* следующим образом. До получения сообщения $x \in X$ от *Tracer User* случайно и равновероятно извлекает $o' \in O(x_0)$ и действует на основании предположения о том, что $o' = o^*$. С учетом изложенных выше предположений средний «доход» *User* равен

$$V(x_0) = \frac{\alpha}{\left|O(x_0)\right|} + \left(1 - \frac{1}{\left|O(x_0)\right|}\right)\beta.$$

При получении $x \in X^+$ *User*, зная, что $o^* \in O(x) \cap O(x_0)$, случайно и равновероятно извлекает $o' \in O(x) \cap O(x_0)$ и получает следующий средний доход

$$V(x, x_0) = \frac{\alpha}{\left|O(x) \cap O(x_0)\right|} + \left(1 - \frac{1}{\left|O(x) \cap O(x_0)\right|}\right)\beta.$$

Под ценностью информации, заключенной в сообщении $x \in X^+$, при наличии априорных сведений $x_0 \in X$ понимается разность

$$\mathrm{Cost}\left(x/x_0\right) = V(x, x_0) - V(x_0).$$

Таким образом, имеет место следующее утверждение [1]

$$\mathrm{Cost}\left(x/x_0\right) = C_{\alpha\beta}\left(\frac{1}{\left|O(x) \cap O(x_0)\right|} - \frac{1}{\left|O(x_0)\right|}\right), \quad (2)$$

где $C_{\alpha\beta} = \alpha - \beta > 0$.

Другими словами, ценность информации пропорциональна увеличению вероятности достижения пользователем цели при использовании этой информации и заключена в следующих границах

$$0 \le \mathrm{Cost}\left(x/x_0\right) \le C_{\alpha\beta}\left(1 - \frac{1}{\left|O(x_0)\right|}\right). \quad (3)$$

Причем $\mathrm{Cost}\left(x/x_0\right) = 0$ в том и только том случае, если $O(x_0) \subseteq (x)$; $\mathrm{Cost}\left(x/x_0\right) = C_{\alpha\beta}\left(1 - \frac{1}{\left|O(x_0)\right|}\right)$, если $\left|O(x) \cap O(x_0)\right| = 1$.

Значение ценности информации, определяемое выражениями (2) и (3), не учитывает фактор времени. Очевидно, что по истечении какого-то пе-

риода времени, ценность информации изменяется (как в сторону уменьшения, так и в сторону увеличения), т.е. изменяется ее актуальность. Актуальность информации изменятся прежде всего вследствие того, что информация о наблюдаемом объекте в прошедшие моменты времени уже не соответствует тому же объему информации, но в некоторый другой момент времени, т.е. информация перестает в ряде случаев быть *достоверной*.

Введем понятие достоверности информации сообщения x: $d(x) = P\{o^* \in O(x)\}$, $x \in X$ и заменим выражение (1) следующим:

$$P\{o = o^* / x\} = \begin{cases} \dfrac{d(x)}{|O(x)|}, & o \in O(x); \\ \dfrac{1-d(x)}{|O| - |O(x)|}, & o \notin O(x), \ x \in X. \end{cases}$$

Г.П. Шанкиным показано, что при полной априорной неопределенности *User* количество информации, заключенной в сообщении $x \in X$, и ее ценность связаны соотношением

$$\mathrm{Cost}^d_{x_0}(x) = \frac{C_{\alpha\beta} d(x)}{|O(x)|}\left[1 - 2^{-H^d_{x_0}(x)}\left(\frac{d(x)}{d(\bar{x})}\frac{|O(x)|}{O(\bar{x})}\right)\right],$$

где $\bar{x}$ – величина, сопряженная с x, а $H^d_{x_0}(x)$ – количество информации, заключенной в сообщении $x \in X$, имеющего достоверность $d(x)$, равно

$$H^d_{x_0}(x) = \log_2 \frac{|O|}{\left(\dfrac{|O(x)|}{d(x)}\right)^{d(x)} \left(\dfrac{|O(\bar{x})|}{d(\bar{x})}\right)^{d(\bar{x})}}.$$

При $d(x) = 1$

$$H_{x_0}(x) = \log_2 \frac{|O|}{|O(x)|}, \ \mathrm{Cost}_{x_0} = \frac{C_{\alpha\beta}}{|O(x)|}\left(1 - 2^{H_{x_0}(x)}\right).$$

Наблюдаемый объект изменяет свои состояния во времени. В некоторый исходный момент времени объект находится в состоянии o^*, которое принадлежит $O(x)$ с вероятностью $d(x_0)$. По истечении некоторого периода времени o^* «выходит» из $O(x)$, и величина $d(x)$ становится равной нулю. Предположим, что смена состояний объекта происходит в дискретные моменты времени с интервалом τ. Тогда за время τ объект с вероятностью $(1 - p)$ остается в том же самом состоянии, а с вероятностью p переходит в любое другое состояние (величина p характеризует динамизм развития объекта). Вероятность невыхода истинного состояния объекта из множества $O(x)$ за время t определяется следующим образом:

$$P_H(t) = \left[(1-p) + p\left(\frac{O(x)-1}{|O|-1} \right) \right]^t, \; t = 1, 2, \ldots,$$

где выражение в квадратных скобках есть вероятность невыхода o^* из $O(x)$ за время τ, а время t изменяется в единицах τ.

Пусть $d_t(x)$ – достоверность сообщения $x \in X$ в момент времени t. Информация появляется в момент времени $t = 0$. Вероятность $P_H(0) = d_0(x)$. При первом же выходе истинного состояния из множества $O(x)$ информация полностью обесценивается. При этом

$$d_t(x) = d_0(x) P_H(t).$$

Ценность $x \in X$ становится равной нулю, когда $d_t(x) = \frac{|O(x)|}{|O|}$.

Обозначим $\lambda_d(x) = \ln\left(1 - p\frac{|O| - |O(x)|}{|O|-1}\right)^{-1} \geq 0$, причем $\lambda_d(x) = 0$ тогда и только тогда, когда $O(x) = O$ и $p > 0$. Тогда $d_t(x) = d_0(x) e^{-\lambda_d(x)t}$, $x \in X$, $t = 0, 1, 2, \ldots$. Величина $\lambda_d(x)$ характеризует интенсивность убывания $d(x)$, эта величина убывает с ростом $|O(x)|$. Точность сообщения $x \in X$ характеризуется величиной $|O(x)|^{-1}$. Поэтому справедливо следующее утверждение: чем выше точность сообщения $x \in X$, тем быстрее во времени убывает его достоверность. В итоге получим

$$\mathrm{Cost}^d_{x_0}(x, t) = C_{\alpha\beta}\left(\frac{d_0(x) e^{-\lambda_d(x)t}}{|O(x)|} - \frac{1}{O} \right). \qquad (4)$$

Обозначим $t^d_{\text{кр}}(x)$ – время, за которое сообщение $x \in X$ полностью обесценится, т.е.

$$\mathrm{Cost}^d_{x_0}\left(x, t^d_{\text{кр}}(x)\right) = 0.$$

Величина $t^d_{\text{кр}}(x)$ определяется из условия

$$\frac{d_0(x) e^{-\lambda_d(x) t^d_{\text{кр}}(x)}}{|O(x)|} \leq \frac{1}{|O|}.$$

Из (4) получим

$$t^d_{\text{кр}}(x) = \left] \frac{1}{\lambda_d(x)} \ln \frac{d_0(x)|O|}{|O(x)|} \right[,$$

где $]a[$ – ближайшее целое, не меньше a. Отсюда следует, что при $t > t^d_{\text{кр}}(x)$ имеет место дезинформация *User* и $\text{Cost}^d_{x_0}(x,t) < 0$. Поэтому $t^d_{\text{кр}}(x)$ является *временем актуальности* сообщения *x*. По истечении данного времени требования к защите деловой информации существенно снижаются и в большинстве случаев вообще не требуют защиты.

Для определения $t^d_{\text{кр}}(x)$ в практических приложениях прежде всего требуется иметь возможность оценить значение величины *p*, от которого зависит величина $\lambda_d(x)$. Значения мощностей множеств $|O|$ и $|O(x)|$ можно с достаточной степенью точности косвенно определить с помощью технических средств (например, оценка объемов информации в дисковых массивах информационных систем). А величину *p*, по мнению автора, следует оценивать, используя аппарат нечеткой логики. Напомним, что *p* характеризует динамизм *изменения* объекта, при этом под динамизмом понимается любое изменение состояния объекта – это может быть как эволюция, так и деградация.

Если возникает необходимость управлять объектом, который обладает неоднозначными свойствами, описание которого заведомо неполно либо не может быть сведено к простой математической модели, то в качестве альтернативного метода управления могут быть выбраны модули нечетко-нейронного управления [2]. Модуль нечеткого управления значением *p* может быть представлен в форме многослойной нейронной сети с прямым распространением сигнала (*feedforward*). Чтобы определить конструкцию модуля, требуется задать его типовые составляющие: базу правил (лингвистическую модель), блок вывода, блок фуззификации и блок дефуззификации. Подробное рассмотрение построения модуля нечеткого управления значением *p* выходит за рамки данной статьи. Приведем лишь окончательные выражения для составляющих модуля.

База правил

$$R^k : IF\left(z_1 \text{ это } A_1^k \text{AND} \ldots \text{AND}\, z_n \text{ это } A_n^k\right)\text{THEN}\left(p \text{ это } B^k\right),$$

где R^k – *k*-ое нечеткое правило ($k = \overline{1, N}$); A_i^j – нечеткие множества входных лингвистических переменных z_i, влияющих на *p*; B^j – нечеткие множества выходной переменной *p*.

Блок вывода

$$\mu_{\bar{B}^k}(p) = \sup_{z_1 \ldots z_n \in Z}\left\{\mu_{B^k}(p)\prod_{i=1}^{n}\mu_{A_i'}(z_i)\mu_{A_i^k}(z_i)\right\},$$

где $\mu(*)$ – функции принадлежности лингвистических переменных.

Блок фуззификации с операцией типа синглетон

$$A'(z) = \begin{cases} 1, \text{ если } z = \bar{z}; \\ 0, \text{ если } z \neq \bar{z}. \end{cases}$$

Блок дефуззификации на основе метода *center average defuzzification*

$$\bar{p} = \frac{\sum_{k=1}^{N} \bar{p}^k \mu_{\bar{B}^k}\left(\bar{p}^k\right)}{\sum_{k=1}^{N} \mu_{\bar{B}^k}\left(\bar{p}^k\right)},$$

где $\bar{p}^k$ – это центр (*center*) нечеткого множества B^k, т.е. точка, в которой $\mu_{B^k}(p)$ достигает максимального значения.

Функция принадлежности гауссовского типа

$$\mu_{A_i^k}(z_i) = \exp\left[-\left(\frac{z_i - \bar{z}_i^k}{\sigma_i^k}\right)^2\right],$$

где $\bar{z}_i^k$ – это центр, а σ_i^k – ширина гауссовской кривой.

В итоге модуль нечеткого управления имеет следующий вид

$$\bar{p} = \frac{\sum_{k=1}^{N} \bar{p}^k \left(\prod_{i=1}^{n} \exp\left[-\left(\frac{\bar{z}_i - \bar{z}_i^k}{\sigma_i^k}\right)^2\right]\right)}{\sum_{k=1}^{N} \left(\prod_{i=1}^{n} \exp\left[-\left(\frac{\bar{z}_i - \bar{z}_i^k}{\sigma_i^k}\right)^2\right]\right)}.$$

Каждый элемент последней формулы может быть задан, как набор функциональных блоков, размещенный в соответствующем слое нейронной сети. В результате получаем четырехслойную гибридную систему с нечетко-нейронным управлением для оценки динамизма изменения состояния наблюдаемого объекта.

Список литературы

1 Шанкин, Г.П. Ценность информации. Вопросы теории и приложений / Г.П. Шанкин. – М.: Филоматис, 2004. – 128 с.

2 Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы : пер. с польск. И.Д. Рудинского / Д. Рутковская, М. Пилиньский, Л. Рутковский. – М.: Горячая линия – Телеком, 2004. – 452 с.

Estimation of Relevance of Protected Information

S.G. Fomicheva

Norilsk Industrial Institute

Key words and phrases: information reliability; information protection; intellectual property; information-analytical systems; value of information.

Abstract: The attempt to formalize the relevance of information and its protection is made. The solution to the problem of intellectual property is in its infancy. It is obvious that it is impossible to protect the whole volume of information. No doubt that information protection requires extra resources in terms of time and money.