

## УПРАВЛІННЯ ІНФОРМАЦІЙНОЮ БЕЗПЕКОЮ

# О МОДЕЛЯХ ПРАГМАТИЧЕСКОЙ ЦЕННОСТИ ИНФОРМАЦИИ

Александр Архипов

Національний технічний університет України "Київський політехнічний інститут"



АРХИПОВ Олександр Євгенійович, д.т.н., професор

Рік та місце народження: 1948 рік, м. Київ, Україна.

Освіта: Київський політехнічний інститут (Національний технічний університет України "Київський політехнічний інститут"), 1972 рік.

Посада: професор кафедри інформаційної безпеки з 1999 року.

Наукові інтереси: інформаційна безпека, математичні моделі систем та процесів, прикладна статистика.

Публікації: більше 200 наукових публікацій, серед яких монографії, навчальні посібники, наукові статті.

E-mail: [sonet@zeos.net](mailto:sonet@zeos.net)

**Аннотация.** Рассмотрена задача определения зависимости ценности информации (в понимании ее полезности для решения некоторой проблемы) от объема этой информации. Введен ряд допущений и ограничений, позволивших формализовать рассмотрение задачи и получить ее решение, представленное в виде математической модели.

**Ключевые слова:** защита информации, ценность информации, полезность информации, модель ценности информации, энтропия.

### Вступлення

Рациональный подход к решению проблемы защиты информации, в частности к созданию систем защиты информации (СЗИ), в общем случае базируется на сопоставлении затрат на СЗИ и рисков потерь, обусловленных реализацией угроз в отношении информации, подлежащей защите. Узким местом этого подхода обычно является необходимость определения параметров рисков: ценности информации и вероятности реализации угроз. Собранная статистика инцидентов, связанных с попытками преодоления СЗИ, позволяет формировать оценки вероятности возникновения и реализации угроз, тогда как определение ценности информации в каждом случае выполняется сугубо индивидуально. Одним из наименее исследованных аспектов определения ценности информации является выявление характера зависимости ценности информации от ее объема.

Некоторые материалы по этой проблеме приведены в [1-3], однако их изложение характеризуется освещением преимущественно качественных аспектов исследований, общим характером и отсутствием формализации полученных в ходе исследований результатов. Учитывая это, актуальным и интересным направлением исследований представляется формализация и моделирование зависимости ценности информации от ее объема, что и является целью данной статьи.

На сегодняшний день прагматический подход – один из наиболее распространенных подходов к определению ценности информации, в котором ценность информации измеряется уровнем ее полезности, в частности, выгодой, получаемой от использования этой информации в конкретных видах человеческой деятельности. Мерой полезности может быть рост производительности труда, обусловленный вовлечением специфических видов и объемов информации в процессы организаторской, управленческой и производственной деятельности, уменьшение себестоимости производимого продукта (оказываемых услуг) и т.п. Денежное выражение величины получаемого позитивного эффекта (меры полезности) определяет прагматическую оценку  $L$  ценности информации, обусловившей этот эффект. "Снятие" полезной информации (ограничение в доступе к ней) приводит к обратному эффекту и соответствующим денежным потерям  $q$ .

### Постановка задачи

Попытаемся проанализировать некоторые особенности нахождения прагматических оценок ценности информации. Предположим, исследуется совокупность проблемных задач в прикладной сфере деятельности, содержание которых в конечном итоге сводится к принятию решения в соответствующей сфере, причем вероятность получения правильного решения зависит от степени полноты специфических для данного вида деятельности

сведений (информации), привлекаемых для выработки конечного решения. Считаем, что на момент постановки любой из задач полностью отсутствует какая-либо информация по сути их решения, однако при этом доступен для ознакомления и изучения практически неограниченный по объему глобальный информационный ресурс (ИР), включающей среди прочего и полезную информацию, необходимую для получения правильного решения поставленной задачи (далее – релевантную информацию, от *relevant* (англ.) – относящийся к данному вопросу, задаче).

В этих условиях начальная стадия поиска решения сводится к выполнению ряда произвольных обращений к глобальному ИР, целью которых является его изучение, первоначальное ознакомление с содержащейся в нем информацией, получение определенных представлений о ее структуре. Несмотря на то, что объемы перерабатываемого ИР могут быть весьма существенными, реальная полезность извлеченных при этом сведений оказывается невысокой из-за спонтанного несистемного характера поиска, реализуемого наугад, фактически в слепую. Темп накопления релевантной информации в данных условиях весьма низок.

Выполнение описанной ознакомительно-исследовательской стадии позволяет перейти к последующей стадии поиска решения, для которой характерна возможность формирования более осознанной последовательности поисковых действий, ориентированных на целенаправленное построение процесса систематизированного извлечения из глобального ИР, накопления и осмысления объемов релевантной информации, используемой для выработки правильного решения. Особенностью новой стадии процесса поиска решения является то, что по мере увеличения объема релевантной информации ее последующее выделение становится все более прогнозируемым и управляемым, темп прироста этой информации возрастает и в какой-то момент (для определенного объема накопленной полезной информации достигает своего максимума). Дальнейший поиск новой релевантной информации в глобальном ИР становится менее результативным, т.к. плотность релевантной информации в перерабатываемых объемах ИР уменьшается по мере роста объема уже выделенной релевантной информации. Эта стадия поиска решения задачи является финальной: уже накопленный достаточно большой объем релевантной информации обеспечивает высокую (требуемую) вероятность получения правильного решения задачи, позволяя прервать процедуру выявления и извлечения новых фрагментов релевантной информации. Но в принципе этот процесс может быть продолжен до получения некоторого объема  $I_{r \max}$  релевантной информации, гарантирующего правильное решение поставленной задачи, т.е.  $P(I_{r \max}) = 1$ , причем значение  $I_{r \max}$  может оказаться некоторой асимптотой, достижимой при устремлении объема  $I$  в бесконечность:

$I_{r \max} = \lim_{I \rightarrow \infty} I_r(I)$ . Другими словами, выделение полного объема релевантной информации  $I_{r \max}$  требует переработки бесконечно большого объема информации общего характера и значительных ресурсов времени.

#### Понятие прагматической ценности информации

Однако в данном случае интерес представляет тот факт, что выделенному конечному объему релевантной информации  $I_{r \max}$  можно поставить в соответствие величину  $Q_{\max}$  потерь, обусловленных (в рамках рассматриваемой задачи) принятием неправильного решения. Избежать эти потери с вероятностью 1 можно только при наличии полного объема релевантной информации, в чем и состоит "полезность" этого объема. Любой другой объем релевантной информации  $I_r < I_{r \max}$  менее "полезен", т.к. для него вероятность правильного решения будет определяться значением  $P(I_r) < 1$ , причем чем меньше значение  $I_r$ , тем ниже соответствующая вероятность  $P(I_r)$ . Исходя из этого можно ввести понятие прагматической ценности  $L(I)$  общей информации объемом  $I$ , которая будет зависеть от количества содержащейся в этом объеме релевантной информации  $I_r$  и определяться выражением:

$$L(I) = Q_{\max} P(I_r). \quad (1)$$

Очевидно, что для максимального значения прагматической ценности будет справедливым равенство:

$$L_{\max} = L(I_{r \max}) = Q_{\max}, \quad (2)$$

и выражение (1) можно будет переписать в виде:

$$L(I) = L_{\max} P(I_r). \quad (3)$$

В общем случае прагматическая ценность некоторого объема информации – это стоимость совокупных потерь, обусловленных уничтожением, модификацией содержащейся в этом объеме релевантной информации, ее недоступностью или разглашением. Таким образом, ценность информации в рамках прагматического подхода зависит от ее конкретного целевого предназначения, с которым связывается понятие релевантной информации. Поэтому в одном и том же объеме общей исходной информации, в зависимости от характера задач, для решения которых она привлекается, будут находиться различные объемы соответствующих задачам количеств релевантной информации.

#### Модели прагматической ценности информации

Учитывая, что в приведенных выше описаниях понятие объема соответствующей информации оставалось не конкретизированным, определим его количеством информации по

Шеннону (т. е. используем синтаксическую меру информации). Очевидно, что объем релевантной информации  $I_r$  монотонно возрастает с увеличением общего объема  $I$  информации, извлекаемой из глобального ИР для решения поставленной задачи. Зависимость  $I_r(I)$  нелинейна: прирост релевантной информации для исчезающе малых (начальных) и больших ( близких к максимально возможному  $I_{\max}$  ) объемов информации  $I$  стремится к нулю

$$\left. \frac{dI_r}{dI} \right|_{I=0} = 0, \quad \left. \frac{dI_r}{dI} \right|_{I=I_{\max}} = 0, \quad (4)$$

но существенно отличен от 0 в средней части интервала  $(0, I_{\max})$  т.е. зависимость  $I_r(I)$  по своему характеру близка к так называемым логистическим (либо сигмоидальным).

К сожалению, а приведенных выше формулах (1) – (4) измеряемым является только количество информации  $I$ , возможность прямого измерения объемов релевантной информации отсутствует. Кроме того, реально определение значения  $L_{\max}$  и получение качественных оценок зависимости  $L(I)$ , позволяющих оценить примерный вид и получить общие представления о ее структуре, что позволяет предположить возможность практического построения приближенной модели этой зависимости.

Отметим, также, что принятое выше в разделе **Постановка задачи** допущение об отсутствии на момент начала решения прикладной проблемы каких-либо сведений по сути ее решения является весьма условным, т.к. сам факт формулирования задания на проведение исследований подразумевает наличие определенной начальной релевантной информации  $I_{r0}$  и соответствующего еще большего объема общей информации  $I_0$ . Учитывая, что производная  $dL/dI$  положительна, и предполагая возможность ее аппроксимации квадратичной параболой вида

$$\frac{dL}{dI} = L(\beta_0 - \beta_1 I); \quad \beta_0, \beta_1 > 0, \quad (5)$$

где  $\beta_0, \beta_1$  – некоторые подлежащие заданию коэффициенты, то, с учетом начальных условий (4), решение дифференциального уравнения (5) приводит к известному соотношению [4], называемому моделью Верхалста-Перла, несколько модифицированной для учета особенностей измерения переменной  $I$ :

$$L(I) = \frac{L_0 \beta_0}{(\beta_0 - L_0 \beta_1) e^{-\beta_0(I-I_0)} + L_0 \beta_1}, \quad I \geq I_0, \quad (6)$$

где  $L_0$  – прагматическая ценность начального объема общей информации  $I_0$ , определяемая

количеством содержащейся в нем релевантной информации  $I_{r0}$ .

К сожалению, приведенные выше качественные и формализованные изложения сценария образования прагматической ценности информации и построения соответствующих моделей адекватны лишь в том случае, когда глобальный ИР функционирует как дискретный источник информации, хранящаяся в котором информация представлена множеством дискретных информационных элементов, обеспечивающих возможность формирования конечных объемов информации (в частности объема  $I_{\max}$ ). Однако обычно извлекаемая из глобального ИР информация представляется множеством элементов непрерывной протяженности. В этом случае получаемая количественная оценка информации не является неким количественным показателем абсолютного уровня информированности, а лишь определяет степень уменьшения априорной неопределенности об исследуемом объекте (решаемой задаче) за счет получения некоторого объема информации  $I$ .

Предположим, что некоторому уровню априорной неопределенности, характеризуемому энтропией  $H_1$ , соответствует крайне низкая (близкая 0) вероятность принятия правильного решения  $P_1$ , тогда как остаточной неопределенности, характеризуемой энтропией  $H_2$  (очевидно, что  $H_1 > H_2$ ), соответствует очень высокое значение вероятности  $P_2$  (правильное решение будет принято почти наверное). Уменьшение энтропии  $\Delta H = H_1 - H_2$  обусловлено приращением общего объема информации на величину  $I = \Delta H$ , причем дальнейшее снятие неопределенности ведет к чрезвычайно медленному росту вероятности ( $P$  асимптотически приближается к 1). В этом случае можно считать, что ценность релевантной информации  $I_r$ , извлеченной из объема общей информации  $I$ , определяется значением  $L_{\max} = Q_{\max}$ . Заметим, что в случае полного снятия неопределенности  $H_2 = 0$ , т.е.  $I = I_{\max} = H_1 = \infty$ , что исключает возможность конечного количественного представления объема  $I_{\max}$ .

Оценим, что влияет на значения энтропий  $H_1, H_2$ . Априорная неопределенность, характеризуемая энтропией  $H_1$ , зависит от целого ряда весьма субъективных факторов, таких как уровень общих знаний лиц, непосредственно участвующих в решении поставленной задачи, имеющегося у них опыта работы в соответствующей проблемной области, личных интеллектуально-психологических данных и т.п., т.е. значение энтропии  $H_1$  может варьироваться в весьма широком диапазоне. Финальная остаточная энтропия  $H_2$ , при которой вероятность принятия правильного решения достаточно высока, обычно зависит от ряда общих объективных причин (общая для всех предельно достижимая точность исходных

сведений и данных, типовые информационно-аналитические технологии и методы обработки данных, т.п.), поэтому в значительно меньшей степени подвержена влиянию индивидуальных качеств лиц, участвующих в решении задачи. Как следствие, диапазон возможных значений энтропии  $H_2$  гораздо уже, чем для  $H_1$ . При этом значения  $I_r$  существенно зависят от априорных (начальных) уровней неопределенности  $H_1$  и не могут выступать в качестве объективной характеристики при анализе объемов полезной информации, используемой для принятия решения. В связи с этим при применении прагматического подхода к оцениванию ценности информации желательно зеркально изменить направленность событий в изложенном выше сценарии, развивая его от достаточно однозначно оцениваемого минимального уровня неопределенности  $H_2$  в сторону значений априорных неопределенностей  $H_1$  путем приращения энтропии  $\Delta H$  (или, что то же самое, введения потерь релевантной информации  $I_r$ ), обуславливающих рост общей неопределенности  $H$ .

Получаемые для этого сценария зависимости  $I_r(H)$ ,  $L(H)$  по своему характеру будут близки рассмотренным выше зависимостям  $I_r(I)$ ,  $L(I)$ .

Для подтверждения этого рассмотрим в качестве модельного примера нахождение прагматической ценности совокупной информации, поступающей в систему управления (СУ) некоторого технологического процесса (ТП). Полагаем, что управление ТП осуществляется на базе совокупности входной информации о характеристиках исходного сырья и параметров, определяющих протекание ТП (например, состав и физико-химические характеристики сырья, технологические параметры процесса: температура, давление и т.п.). Недостаточность или искажение информации на любом этапе работы СУ (погрешности во входной информации, ошибки оператора, контролирующего состояние ТП, недостатки программно-алгоритмического обеспечения системы поддержки принятия решения оператором) ведут к нарушениям нормального течения ТП и, соответственно, возникновению производственных потерь  $q$ , обусловленных появлением брака, или даже к возможной остановке производства. Последнему случаю соответствует максимально возможное значение  $Q_{\max}$  потерь  $q$ , определяющее прагматическую ценность  $L_{\max}$  релевантной информации, поступающей в СУ.

Обозначим объем информации, обеспечивающий нормальное функционирование СУ и гарантирующий требуемые характеристики качества продукта на выходе ТП (т.е. случай, когда  $q=0$ ) через  $I_{\max}$ , а объем информации, реально поступающий в СУ, – через  $I$ . Очевидно, что в общем случае величина потерь  $q$  зависит от уровня существующей в СУ информационной

неопределенности, характеризуемой приращением энтропии  $\Delta H = I_{\max} - I$ , т.е.  $q$  является функцией от  $H$ .

Предположим, что в данной системе осуществляется постепенное сокращение объема информации (начиная от ее максимально полного значения  $I_{\max}$ ), обеспечивающей работу СУ. Первоначально это сокращение практически не сказывается на работе СУ (а значит и на качестве конечного продукта, получаемого на выходе ТП), так как необходимые для функционирования СУ сведения и данные достаточно полно восстанавливаются за счет наличия информационной избыточности в еще доступных фрагментах исходной информации, существования структурных связей между этими фрагментами, семантической избыточности обрабатываемой в СУ информации. Но последующее сокращение общего конце-концов ведет к падению эффективности компенсаторно-восстановительных механизмов, функционирование которых возможно только в условиях наличия всевозможных видов информационной избыточности. В этой ситуации каждое новое, даже незначительное уменьшение объема поступающей в СУ информации влечет за собой все возрастающие потери качества управления ТП, в частности, оператор начинает испытывать все больший дефицит необходимых для осуществления осмысленных управленческих действий сведений и вынужден все чаще прибегать к принятию интуитивных решений. Дальнейшее сокращение объема поступающей в СУ все меньше влияет на качество управления ТП, так как в данной ситуации управление оказывается практически полностью зависимым от опыта и интуиции оператора, уровня его базовых знаний и интеллекта, психологических и волевых качеств.

Прекращение производственного процесса определяет максимально возможный уровень потерь  $Q_{\max}$ , обусловленный непоступлением информации в СУ. Хотя остаточная информация  $I_0$  отличается от нуля, ее не достаточно для возобновления ТП, т.е. прагматическая ценность  $L_{\max}$  всей информации  $I_{\max}$ , необходимой для полноценной работы СУ, соответствует значению максимально возможных потерь  $Q_{\max}$ . Тогда, как следует из приведенного описания (сценария), приращение уровня потерь  $q(H)$ , обусловленных сокращением объема информации, поступающей в СУ, определяет значение  $L(H) = L_{\max} - q(H)$  прагматической ценности текущей информации  $I$ , т.е. прагматическая ценность информации полностью определяется через характеристики потерь информации или, что то же самое, приращение энтропии:  $\Delta H = I_{\max} - I$ . Очевидно, что характеристика  $L(H)$  – монотонно убывающая. Кроме того, приращения информации в окрестностях значений информации  $I_{\max}$  (окрестностях значения энтропии  $H_2$ ) практически не сказывается на управляемости ТП, т.е.

прагматическая ценность этих приращений фактически равна нулю. То же самое справедливо относительно приращений энтропии  $H$  в окрестности точки  $H_1$ . В итоге получаем:

$$\left. \frac{dL}{dH} \right|_{H_2} = 0, \quad \left. \frac{dL}{dH} \right| < 0, \quad \text{при } H \in (0, H_{\max}). \quad (7)$$

Учитывая нелинейный характер зависимости  $L(H)$ , предполагаем, что производная  $dL/dH$  может быть аппроксимирована квадратичной функцией вида:

$$\frac{dL}{dH} = -L(b_0 - b_1 L); \quad b_0, b_1 > 0. \quad (8)$$

Соотношение (8) - дифференциальное уравнение с разделяющимися переменными, решая которое, получаем:

$$\ln \frac{AL}{b_0 - b_1 L} = -b_0 H, \quad (9)$$

где  $A$  - постоянная интегрирования. Потенцирование (9) приводит к соотношению

$$\frac{AL}{b_0 - b_1 L} e^{-b_0 H}. \quad (10)$$

Учитывая граничное условие  $L(0) = L_{\max}$ , для  $H = 0$  из соотношения (10) получаем:

$$A = \frac{b_0}{L_{\max}} - b_1 = \frac{1}{L_{\max}} (b_0 - b_1 L_{\max}), \quad (11)$$

что после подстановки в уравнение (10) дает

$$L(H) = \frac{b_0}{b_1 + \left( \frac{b_0}{L_{\max}} - b_1 \right) e^{b_0 H}}. \quad (12)$$

Формула (6) дает аналитическое выражение зависимости прагматической ценности  $L$  информации от вносимой в СУ энтропии  $H$ . Определяя энтропию как разность  $H = I_{\max} - I$  и подставляя это выражение в формулу (12), после проведения простых преобразований находим

зависимость прагматической ценности информации от ее объема:

$$L(I) = \frac{L_{\max} b_0}{(b_0 - L_{\max} b_1) B e^{-b_0 I} + L_{\max} b_1}, \quad (13)$$

где  $B = e^{b_0 I_{\max}}$ .

**Выводы.** В статье дано определение понятия "прагматическая ценность информации", приведены модели (6), (13) зависимости ценности информации от ее объема и модель (12) зависимости ценности информации от уровня вносимой неопределенности. Модель (6) ориентирована на оценивание ценности информации в стадии ее начального накопления и при работе со средними объемами накопленной релевантной информации, модель (12) - на работу со средними и большими объемами этой информации, модель (13) позволяет оценить стоимость потерь информации в исходно полных объемах релевантной информации.

#### Литература

- [1] Сурмин Ю. П. Теория систем и системный анализ / Ю. П. Сурмин - К.: МАУП, 2003. - 368 с.
- [2] Архипов О.Є., Ворожко В.П. Системні аспекти оцінювання рівня важливості секретної інформації. // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні, випуск 2(15).- Київ-2007р.- с. 10-12.
- [3] Архипов О.Є., Муратов О.Є. Критерії визначення можливої шкоди національній безпеці України у разі розголошення інформації, що охороняється державою: моногр. / О.Є.Архипов, О.Є.Муратов. - К.: Наук.-вид. відділ НА СБ України, 2011. - 195с.
- [4] Кучин Б.Л., Якушева Е.В. Управление развитием экономических систем. / Б.Л. Кучин, Е.В.Якушева. - М.: Экономика, 1990. - 157 с.

УДК 004.056.5 (045)

#### Архипов О.Є. Про моделі прагматичної цінності інформації

**Анотація.** Розглянуто задачу визначення залежності цінності інформації (в розумінні її корисності при розв'язанні певної проблеми) від обсягу цієї інформації. Введено деякі припущення та обмеження, що дозволили формалізувати розгляд задачі та отримати її рішення у формі математичної моделі.

**Ключові слова:** захист інформації, цінність інформації, корисність інформації, модель цінності інформації, ентропія.

#### Arkhyrov O.E. About models of the pragmatic information value

**Abstract.** The problem of determination of the information value dependence (in the sense of its usefulness in solving a certain problem) of the information amount is considered. It is introduced some assumptions and limitations that allow to formalize the solving of the problem and obtain its solution in the form of a mathematical model.

**Keyword:** information security, information value, information usefulness, model of information value, entropy.

Отримано 19 квітня 2012 року, затверджено редколегією 05 червня 2012 року  
(рецензент д.т.н., професор О.Г. Корченко)