XXIX научно-практическая конференция «Комплексная защита информации»

НЕКОТОРЫЕ ПОДХОДЫ К ОБОСНОВАНИЮ КРИТЕРИЕВ ЭФФЕКТИВНОСТИ ЗАЩИТЫ РЕЧЕВОЙ И ТЕКСТОВОЙ ИНФОРМАЦИИ ОТ ЕЕ УТЕЧКИ ПО ТЕХНИЧЕСКИМ КАНАЛАМ

Хорев Анатолий Анатольевич

профессор, доктор технических наук

г. Санкт – Петербург, 2024 г.

Показатели и критерии эффективности защиты речевой и текстовой информации от ее утечки по техническим каналам

В качестве **показателя эффективности** защиты речевой (текстовой) информации от ее утечки по техническим каналам наиболее часто используется **разборчивость речи (текста), *И*,** отображающая качественную область понятности перехваченного разговора (текста), под которой понимается отношение количества правильно распознанных слов (фраз) к общему количеству слов (фраз) в перехваченном разговоре (тексте).

Критерии эффективности защиты речевой или тестовой информации во многом зависят от целей, преследуемых при защите информации, например, скрыть тематику ведущегося разговора (текста) или скрыть его смысловое содержание.

Цели защиты информации	Показатели эффективности защиты информации	Критерии эффективности защиты информации	Пороговое значение разборчивости речи (текста)
Скрытие тематики разговора (текста)	Словесная разборчивость речи (текста), $W_{сл}$	Количество правильно распознанных слов не позволяет установить тематику перехваченного разговора (текста), $W_{сл} \le W_{сл.п}$	**$W_{сл.п} = 0,2$**
Скрытие содержания разговора (текста)	Словесная разборчивость речи (текста), $W_{сл}$	Количество правильно распознанных слов не позволяет составить аннотацию (краткую справку) о перехваченном разговоре (тексте), $W_{сл} \le W_{сл.п}$	**$W_{сл.п} = 0,3$**

Однако **не все распознанные слова или фразы относятся к ключевым**, по которым можно установить тематику перехваченного разговора (текста) и составить его аннотацию.

Следовательно, для оценки эффективности защиты речевой (текстовой) информации необходимо не только рассчитать словесную или фразовую разборчивость речи (текста), но и оценить количество ключевых слов и фраз, требуемых для определения тематики разговора (текста) или составления его аннотации.

При этом целесообразно использовать вероятностный метод оценки вскрытия тематики перехваченного разговора (текста) и составления его аннотации.

Будем полагать, что **тематика разговора (текста) вскрыта**, если количество распознанных ключевых слов $N_{сл.кл.р}$ будет не менее установленного порогового значения $N_{сл.кл.min}$, а **аннотация разговора составлена**, если количество распознанных ключевых фраз $N_{фр.кл.р}$ будет не менее установленного порогового значения $N_{фр.кл.min}$.

Учитывая, что средний темп речи на русском языке 120 –130 слов в минуту, за 10 минут можно в среднем произнести 1200 – 1300 или 140 – 180 фраз.

Для каждого фрагмента разговора (текста) необходимо определить его тематику и составить аннотацию.
Будем полагать, что цели защиты речевой информации достигнуты, если для каждого фрагмента разговора нельзя соответственно определить его тематику, или нельзя составить его аннотацию.

Вероятность вскрытия тематики разговора (текста) P_m, то есть вероятность того, что количество распознанных ключевых слов будет не менее установленного значения ($N_{сл.кл.р} \geq N_{сл.кл.min}$)

$$P_m\left(N_{\text{сл.кл.р}} \geq N_{\text{сл.кл.}min}\right) = \sum_{i=N_{\text{сл.кл.}min}}^{N_{\text{сл.кл.р}}} P_{\text{т}}\left(N_{\text{сл.кл.р.}i}\right); \quad (1) \qquad P_{\text{т}}\left(N_{\text{сл.кл.р.}i}\right) = \frac{C_{N_{\text{сл.кл}}}^{N_{\text{сл.кл.р.}i}} \cdot C_{N_{\text{сл}}-N_{\text{сл.кл}}}^{N_{\text{сл.кл}}-N_{\text{сл.кл.р.}i}}}{C_{N_{\text{сл}}}^{N_{\text{сл.кл}}}}; \qquad (2)$$

где $C_b^a = \frac{b!}{a!\cdot(b-a)!}$ – формула комбинаторики, определяющая количество сочетаний без повторений [3];

$N_{\text{сл}}$ – количество слов в перехваченном разговоре;

$N_{\text{сл.кл}} = k_{\text{сл}} \cdot N_{\text{сл}}$ – количество ключевых слов в перехваченном разговоре;

$k_{\text{сл}}$ – среднее относительное количество ключевых слов по данной тематике;

$N_{\text{сл.р}} = W_{\text{сл}} \cdot N_{\text{сл}}$ – количество распознанных слов в перехваченном разговоре;

$W_{\text{сл}}$ – словесная разборчивости перехваченного разговора;

$N_{\text{сл.кл.р}} = W_{\text{сл}} \cdot N_{\text{сл.кл}} = W_{\text{сл}} \cdot k_{\text{сл}} \cdot N_{\text{сл.р}}$ – количество распознанных ключевых слов в перехваченном разговоре;

$N_{\text{сл.кл.}min}$ – минимальное количество ключевых слов, необходимых для вскрытия тематики перехваченного разговора.

Учитывая, что $b>a$ формулу комбинаторики запишем в виде: $C_b^a = \frac{b!}{a!\cdot(b-a)!} = \frac{(a+1)\cdot(a+2)\cdot(a+3)\cdots b}{(b-a)!} = \frac{\prod_{j=a+1}^{b} j}{\prod_{j=1}^{b-a} j}$. (3)

Подставляя (3) в (2) получим: $P_m\left(N_{\text{сл.кл.р.}i}\right) = \frac{\prod_{j=N_{\text{сл.кл.р.}i}+1}^{N_{\text{сл.кл}}} j \cdot \prod_{j=N_{\text{сл.кл}}-N_{\text{сл.кл.р.}i}+1}^{N_{\text{сл}}-N_{\text{сл.кл}}} j \cdot \prod_{j=1}^{N_{\text{сл}}-N_{\text{сл.кл}}} j}{\prod_{j=1}^{N_{\text{сл.кл}}-N_{\text{сл.кл.р.}i}} j \cdot \prod_{j=1}^{(N_{\text{сл}}-N_{\text{сл.кл}})-(N_{\text{сл.кл}}-N_{\text{сл.кл.р.}i})} j \cdot \prod_{j=N_{\text{сл.кл}}+1}^{N_{\text{сл}}} j}$. (4)

Вероятность составления аннотации разговора (текста) $P_{ан}$, то есть вероятность того, что количество распознанных ключевых фраз будет не менее установленного значения ($N_{фр.кл.р} \geq N_{фр.кл.min}$):

$$P_{\text{ан}}\left(N_{\text{фр.кл.р}} \geq N_{\text{фр.кл.}min}\right) = \sum_{i=N_{\text{фр.кл.}min}}^{N_{\text{фр.кл.р}}} P_{\text{т}}\left(N_{\text{фр.кл.р.}i}\right); \quad (5)$$

$$P_{\text{ан}}\left(N_{\text{фр.кл.р.}i}\right) = \frac{\prod_{j=N_{\text{фр.кл.р.}i}+1}^{N_{\text{фр.кл}}} j \cdot \prod_{j=N_{\text{фр.кл}}-N_{\text{фр.кл.р.}i}+1}^{N_{\text{фр}}-N_{\text{фр.кл}}} j \cdot \prod_{j=1}^{N_{\text{фр}}-N_{\text{фр.кл}}} j}{\prod_{j=1}^{N_{\text{фр.кл}}-N_{\text{фр.кл.р.}i}} j \cdot \prod_{j=1}^{(N_{\text{фр}}-N_{\text{фр.кл}})-(N_{\text{фр.кл}}-N_{\text{фр.кл.р.}i})} j \cdot \prod_{j=N_{\text{фр.кл}}+1}^{N_{\text{фр}}} j}. \quad (6)$$

где $N_{фр}$ – количество фраз в перехваченном разговоре;

$N_{фр.кл}$ – количество ключевых фраз в перехваченном разговоре;

$N_{фр.кл.р}$ – количество распознанных ключевых слов в перехваченном разговоре;

$N_{фр.кл.min}$ – минимальное количество ключевых фраз, необходимых для вскрытия тематики перехваченного разговора;

$W_{фр}$ – словесная разборчивости перехваченного разговора.

Методика обоснования критериев эффективности защиты речевой информации от ее утечки по техническим канала

1) Экспериментально определяется относительное среднее количество ключевых слов $k_{сл}$ и фраз $k_{фр}$ по различной тематике.

2) По каждой тематике экспериментально определяются среднее количество ключевых слов $N_{сл.кл}$ и ключевых фраз $N_{фр.кл}$ в разговоре, длительностью 10 минут (теста, объемом 1300 слов).

3) По каждой тематике экспериментально определяются минимальные количество ключевых слов $N_{сл.кл.min}$, необходимых для вскрытия тематики разговора длительностью 10 минут (теста, объемом 1300 слов), и минимальное количество ключевых фраз $N_{фр.кл.min}$, необходимых для составления его аннотации.

4) Экспертным методом определяются пороговые значения вероятностей вскрытия тематики перехваленного разговора $P_{т.п}$, длительностью 10 минут (теста, объемом 1300 слов), и вероятности составления аннотации его аннотации $P_{ан.п.}$

5) По формулам (1) и (2) строится зависимость вероятности вскрытия тематики перехваленного разговора $P_{т}$, длительностью 10 минут (теста, объемом 1300 слов), от словесной разборчивости речи $W_{сл.}$.

6) По графику зависимости вероятности вскрытия тематики перехваленного разговора $P_{т}$ от словесной разборчивости речи $W_{сл}$ для порогового значения вероятности вскрытия тематики перехваленного разговора (текста) $P_{т.п}$ определяется пороговое значение словесной разборчивости речи $W_{сл.п}$.

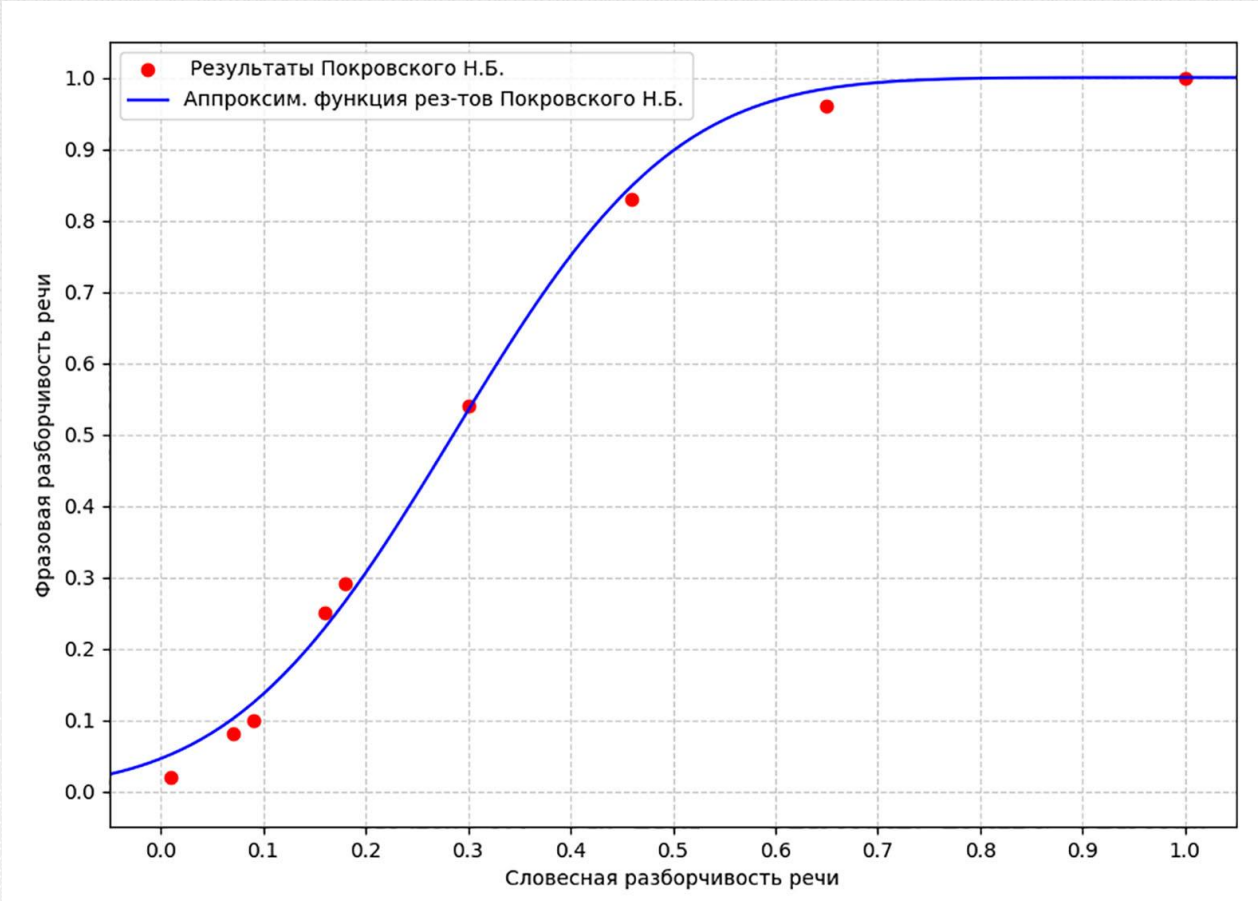
7) По формулам (5) и (6) строится зависимость вероятности составления аннотации разговора (теста) $P_{ан}$, длительностью 10 минут (теста, объемом 1300 слов), от фразовой разборчивости речи $W_{фр}$.

8) По графику зависимости вероятности составления аннотации разговора (текста) $P_{ан}$, длительностью 10 минут (теста, объемом 1300 слов), от фразовой разборчивости речи $W_{фр}$ для порогового значения вероятности составления аннотации разговора $P_{ан.п}$ определяется пороговое значение фразовой разборчивости речи (текста) $W_{фр.п}$.

Среднее количество ключевых слов и ключевых фраз зависит от тематики разговора.

Например, по результатам экспериментальных исследований установлено, что:

- в докладах по тематике «информационная безопасность» среднее значение ключевых слов составляло 2,4%, а среднее значение ключевых фраз – 10,3% ;
- в текстах по научной тематике среднее количество ключевых фраз составило 10,9%;
- в текстах различной направленности среднее количество ключевых слов составляет от 3,2%, а среднее количество ключевых фраз – 11,8%.



Зависимость фразовой разборчивости речи ($W_{фр}$) от словесной ($W_{сл}$)

$$W_{\text{фр}} \approx \Phi(5{,}91 \cdot W_{\text{сл}} - 1{,}69), \quad (1)$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} exp\left(-\frac{t^2}{2}\right) dt$ – интеграл вероятности.

$$W_{\text{сл}} \approx 0{,}169 \cdot \left[\Phi^{-1}\left(W_{\text{фр}}\right) + 1{,}69\right], \quad (2)$$

где $\Phi^{-1}(x)$ – функция, обратная интегралу вероятности.

$W_{сл}$	Вероятность вскрытия тематики разговора P_m				$W_{фр}$	Вероятность составления аннотации разговора $P_{ан}$			
	$N_{сл.кл.р} \geq 4$	$N_{сл.кл.р} \geq 5$	$N_{сл.кл.р} \geq 6$	$N_{сл.кл.р} \geq 7$		$N_{фр.кл.р} \geq 7$	$N_{фр.кл.р} \geq 8$	$N_{фр.кл.р} \geq 9$	$N_{фр.кл.р} \geq 10$
0,10	0,400	0,209	0,092	0,034	0,10	0,000	0,000	0,000	0,000
0,15	0,731	0,537	0,345	0,192	0,15	0,003	0,000	0,000	0,000
0,20	0,910	0,799	0,642	0,465	0,20	0,020	0,004	0,001	0,000
0,25	0,976	0,933	0,850	0,725	0,25	0,069	0,021	0,005	0,001
0,30	0,995	0,982	0,951	0,890	0,30	0,164	0,064	0,020	0,005
0,35	0,999	0,996	0,987	0,965	0,35	0,304	0,147	0,057	0,017
0,40	1,000	0,999	0,997	0,992	0,40	0,473	0,274	0,130	0,049
0,45	1,000	1,000	1,000	0,998	0,45	0,642	0,434	0,245	0,112
0,50	1,000	1,000	1,000	1,000	0,50	0,785	0,603	0,397	0,215
0,55	1,000	1,000	1,000	1,000	0,55	0,888	0,755	0,566	0,358
0,60	1,000	1,000	1,000	1,000	0,60	0,951	0,870	0,726	0,527
0,65	1,000	1,000	1,000	1,000	0,65	0,983	0,943	0,853	0,696
0,70	1,000	1,000	1,000	1,000	0,70	0,995	0,980	0,936	0,836
0,75	1,000	1,000	1,000	1,000	0,75	0,999	0,995	0,979	0,931
0,8	1,000	1,000	1,000	1,000	0,80	1,000	0,999	0,996	0,980
0,85	1,000	1,000	1,000	1,000	0,85	1,000	1,000	1,000	0,997
0,90	1,000	1,000	1,000	1,000	0,90	1,000	1,000	1,000	1,000
0,95	1,000	1,000	1,000	1,000	0,95	1,000	1,000	1,000	1,000
1,00	1,000	1,000	1,000	1,000	1,00	1,000	1,000	1,000	1,000

Разговор, длительностью 10 минут, состоящий из 1300 слов (2,5% из которых ключевые) и 160 фраз (10% из которых ключевые).

Зависимости вероятностей скрытия тематики разговора (P_m) и составления его аннотации ($P_{ан}$) от словесной ($W_{сл}$) и фразовой ($W_{фр}$) разборчивости речи от распознанных ключевых слов ($N_{сл.кл.р}$) и ключевых фраз ($N_{фр.кл.р}$)

Разговор, длительностью 10 минут, состоящий из 1300 слов (2,5% из которых ключевые) и 160 фраз (10% из которых ключевые) при $N_{сл.кл.min} = 7$, а $N_{фр.кл.min} = 10$.

Показатели и критерии эффективности защиты речевой и текстовой информации от ее утечки по техническим каналам

Цели защиты речевой информации	Показатели эффективности защиты речевой информации	Критерии эффективности защиты речевой и текстовой информации		
		Пороговая вероятность скрытия тематики разговора $P_{т.п}$	Пороговая словесная разборчивость речи $W_{сл.п}$	Пороговая фразовая разборчивость речи $W_{фр.п}$
Скрытие тематики текста	Вероятность скрытия тематики разговора $P_т$	0,10	0,12	–
		0,20	0,15	–
		0,30	0,17	–
		0,40	0,20	–
Скрытие содержания текста	Вероятность составления аннотации разговора $P_{ан}$	0,10	0,26	0,45
		0,20	0,29	0,50
		0,30	0,30	0,53
		0,40	0,31	0,56

Спасибо за внимание