УДК 519.717.7
ББК 22.181
В 58

Власенко А.В.
Кандидат технических наук, доцент кафедры компьютерных технологий и информационной безопасности института информационных технологий и безопасности, начальник управления аспирантуры и докторантуры Кубанского государственного технологического университета, Краснодар, e-mail: Vlasenko@kubstu.ru
Дзьобан П.И.
Аспирант кафедры компьютерных технологий и информационной безопасности института информационных технологий и безопасности Кубанского государственного технологического университета, Краснодар, e-mail: antiemoboy@mail.ru
Жук Р.В.
Аспирант кафедры компьютерных технологий и информационной безопасности института информационных технологий и безопасности Кубанского государственного технологического университета, Краснодар, e-mail: goonerkrd@gmail.com

Защита персональных данных при авторизации пользователя в распределенных информационных системах, построенных на основе Web-технологий

(Рецензирована)

Аннотация. *Рассмотрена проблематика идентификации активов при организации защиты информационных систем, построенных на основе Web-технологий. Основываясь на статистических данных об атаках на входные данные пользователей в сети, предложено описание алгоритма идентификации фишинговых Web-сайтов/ресурсов, реализованного в алгоритме предавторизационной проверки. Данный алгоритм является превентивной мерой, реализован посредством эвристического подхода и является реально действующим дополнением Web-браузера в виде плагина. Таким образом, пользователь до ввода каких либо данных в окне браузера сначала убеждается в легитимности Web-ресурса.*

Ключевые слова: *информационная система, актив, персональные данные, фишинг, аутентификация, байесовская сеть, Web-браузер, условные вероятности.*

Vlasenko A.V.
Candidate of Technical Sciences, Associate Professor of Computer Technologies and Information Security Department, Institute of Information Technologies and Security, Head of Department of Postgraduate and Doctoral Studies, Kuban State University of Technology, Krasnodar, e-mail: Vlasenko@kubstu.ru
Dzoban P.I.
Post-graduate student of Computer Technologies and Information Security Department, Institute of Information Technologies and Security, Kuban State University of Technology, Krasnodar, e-mail: antiemoboy@mail.ru
Zhuk R.V.
Post-graduate student of Computer Technologies and Information Security Department, Institute of Information Technologies and Security, Kuban State University of Technology, Krasnodar, e-mail: goonerkrd@gmail.com

Protection of personal data when authorizing a user in distributed information systems based on Web-technologies

Abstract. *The paper considers the problem of asset identification when organizing the protection of information systems built on the basis of Web technologies. Based on the statistical data on attacks on the input data of users on the network, a description of the algorithm for identifying phishing Web sites/resources implemented in the algorithm for pre-authorization verification was proposed. This algorithm is a preventive measure, implemented through a heuristic approach and is a real-life add-on of a Web browser in the form of a plug-in. Thus, before entering any data in the browser window, the user first becomes convinced of the legitimacy of the Web resource.*

Keywords: *information system, asset, personal data, phishing, authentication, Bayesian network, Web-browser, conditional probabilities.*

По данным исследователей [1], во втором квартале 2016 года был зафиксирован рост фишинговой активности в экономическом секторе Интернета. Злоумышленники наиболее активны при подмене Интернет-сайтов кредитных организаций и Интернет-магазинов, а также различных платежных систем. Наряду с этим возрастает количество атак на рядовых

пользователей при использовании социальных сетей, а также на совершенно новом направлении браузерных «онлайн-игр».

В последние годы в Российской Федерации набирают популярность информационные системы сдачи электронной отчетности, обрабатывающие значительный объем информации, существенная часть которой содержит персональные данные. Основной функционал данных систем построен на Web-технологиях. Одним из ключевых вопросов функционирования данных систем является вопрос информационной безопасности как пользовательской информации, являющейся коммерческой тайной, так и авторизационных данных самих пользователей, передающихся по каналам связи в процессе авторизации в информационной системе.

Необходимо отметить, что в рамках данной статьи для авторов объектом исследования является клиентская часть информационной системы (функционирующая на рабочем месте пользователя). Далее термин «информационная система» будет применяться для описания конечного узла пользователя.

Используя разработанную на основе действующих нормативных документов [2, 3] методику, построим модель угроз информационной безопасности для информационной системы.

Установим границы информационной системы, определив основные и вспомогательные активы.

К основным активам отнесем:

– информацию, а именно персональные данные пользователя (логин, адрес электронной почты, пароль, ФИО, ИНН, наименование организации, номер контактного телефона и должность);

– процесс регистрации/входа в информационную систему (Web-приложение).

К второстепенным активам путем перебора элементов отнесем:

– рабочее место пользователя информационной системы;

– общесистемное программное обеспечение;

– прикладное программное обеспечение (Интернет-браузер).

Ценность вспомогательного актива определим как множество

Ц(А)=(«факт обработки ПДн», «показатель степени возможного вреда»),

где «показатель степени возможного вреда»>2*«среднее»/«высокая».

Исходя из информации об основном активе, установим 4-й уровень защищенности для информационной системы. Соответственно для информационной системы будут актуальны угрозы информационной безопасности третьего типа [3, 4], не связанные с наличием недекларируемых возможностей в программном обеспечении рабочего места.

Таким образом, организация защиты активов будет осуществляться в соответствии с требованиями, описанными в работе [5].

Отсутствие защитных мер [5] обусловлено спецификой применения Web-технологий. В рамках существующих методик защита информации при передаче по каналам связи реализуется посредством технологии VPN. Однако защищенное соединение выполняется после процедуры авторизации на Web-сайте, в связи с чем остается ряд рисков, связанных с угрозами фишинга в процессе авторизации.

Наряду с этим необходимо отметить, что современные Web-браузеры, а также применяемое антивирусное программное обеспечение реализуют механизм защиты от фишинга на основе применения черных и белых списков доступа. Данные списки могут храниться как на компьютере пользователя информационной системы, так и в специально разработанном облаке (часто применяется разработчиками антивирусного программного обеспечения). Механизм функционирования черных и белых списков достаточно прост:

1. При открытии каждого нового соединения адрес открываемого Web-ресурса передается на сервера, где происходит проверка открываемого адреса на наличие в черном списке. Если адрес присутствует в черном списке, Web-браузер отображает соответствующее сообщение об ошибке.

2. При открытии соединения адреса из белых списков считаются доверенными, остальные адреса передаются на сервер с черными списками.

При этом возникает проблема рационального использования ресурсов Web-браузером.

С другой стороны, существенной проблемой является время актуализации вышеперечисленных списков. Несмотря на значительный прогресс, продолжительность реагирования на появление фишингового ресурса велика.

Проблему своевременности реагирования на открытие пользователем фишингового Web-ресурса можно решить путем внедрения в Web-браузер алгоритма предавторизационной проверки Web-ресурсов. Реализация данного алгоритма позволит свести к минимуму время реагирования на фишинговый Web-ресурс.

Для реализации данного алгоритма используется математическая модель на основе байесовской сети. Основной решаемой задачей математической модели является ответ на вопрос, с какой вероятностью Web-ресурс является безопасным для пользователя, который собирается ввести авторизационные данные.

Байесовская сеть работает тогда и только тогда, когда построенный граф соответствует реальности, а условные вероятности вершин графа вычислены из большой выборки исходных данных. Анализ ресурса Open Web Application Security Project (OWASP) и рейтинга OWASP Top 10 [1] позволил выделить следующие критерии для построения математической модели:

- наличие предыдущих посещений Web-ресурса (X_1);
- использование протокола SSL/TLS (X_2);
- наличие похожих доменных имен (X_3);
- наличие смешанных символов в доменном имени (X_4).
- частота смены IP-адреса Web-ресурса (X_5);
- факт смены удостоверяющего центра (X_6);
- факт смены имени сертификата (X_7);
- версия протокола SSL/TLS (X_8);
- срок действия сертификата (X_9);
- наличие HTTP Strict Transport Security (HSTS) (X_{10});
- геопозиция IP-адреса (X_{11});
- удостоверяющий центр сменился на зарубежный (X_{12});
- факт действия предыдущего сертификата (X_{13});
- самоподписанность сертификата (X_{14});
- сертификат некоммерческого удостоверяющего центра (X_{15});
- сертификат коммерческого удостоверяющего центра (X_{16});
- наличие сети взаимодоверяющих удостоверяющих центров (X_{17});

Графическая модель связей критериев представлена на рисунке 1.

Распределение условных вероятностей (их количество 2^{17} и в сумме они дают 1) открытия пользователем фишингового Web-ресурса представлено в следующем виде:

$$P(X_0..X_{17}) = P(X_0)P(X_1|X_0)P(X_2|X_0)P(X_3|X_0)P(X_4|X_0)P(X_5|X_1)P(X_6|X_1,X_{15}X_{16})P(X_7|X_1,X_2)$$
$$P(X_8|X_1,X_2)P(X_9|X_2)P(X_{10}|X_2)P(X_{11}|X_5)P(X_{12}|X_6)P(X_{13}|X_7)P(X_{14}|X_9)P(X_{15}|X_9)P(X_{16}|X_9)$$
$$P(X_{17}|X_{15},X_{16}).$$

В разработанной методике вероятность подмены Web-ресурса вычисляется на основе статистических данных об уникальных фишинговых атаках за первую половину 2017 года в доменной зоне “ru”.

Для каждого критерия «X» разрабатывается таблица с параметрами «истина» и «ложь», количественные показатели которых присвоены на основе статистической вероятности их реализации (табл. 1–18). Величина оценки каждого параметра расположена в диапазоне значений от 0 до 1. Разрабатываемый алгоритм предполагает накопление и хранение статистической информации для корректировки точности параметров каждого из приведенных критериев.

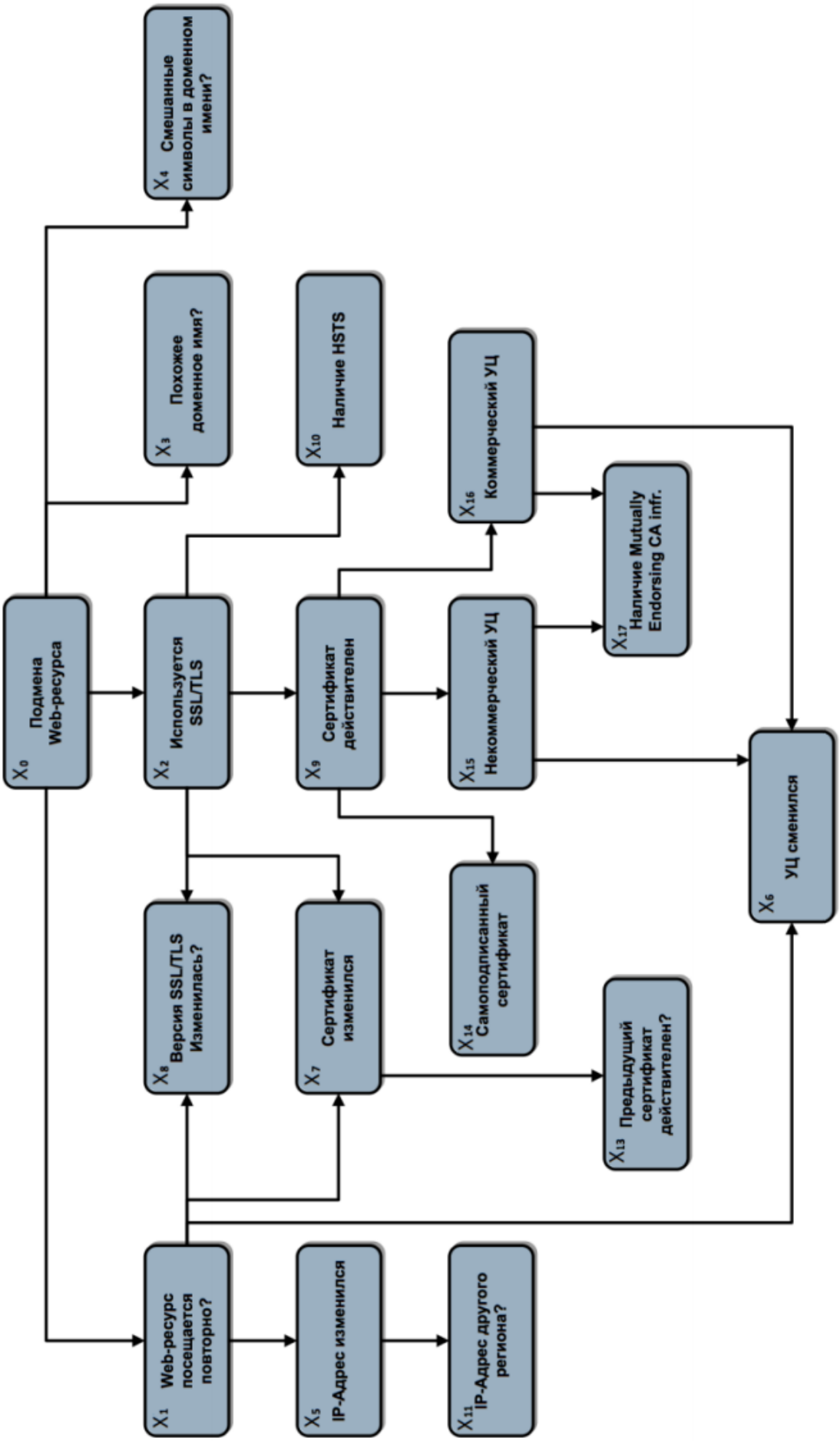

Рис. 1. Графическая модель алгоритма взаимосвязи показателей априорных условных вероятностей

Таблица 1

Критерий X_0 (подмена ресурса)

X_0 (подмена ресурса)	
Истина	Ложь
0,0003	0,9997

Таблица 2

Критерии X_0 (подмена ресурса) и X_1 (повторное посещение)

X_0, подмена ресурса	X_1, повторное посещение	
	Истина	Ложь
Истина	0,96	0,04
Ложь	0,9998	0,0002

Таблица 3

Критерии X_0 (подмена ресурса) и X_2 (используется SSL/TLS)

X_0, подмена ресурса	X_2, используется SSL/TLS	
	Истина	Ложь
Истина	0,2	0,8
Ложь	0,76	0,24

Таблица 4

Критерии X_0 (подмена ресурса) и X_3 (похожее доменное имя)

X_0, подмена ресурса	X_3, похожее доменное имя	
	Истина	Ложь
Истина	0,85	0,15
Ложь	0,97	0,03

Таблица 5

Критерии X_0 (подмена ресурса) и X_4 (смешанные символы в доменном имени)

X_0, подмена ресурса	X_4, смешанные символы в доменном имени	
	Истина	Ложь
Истина	0,00002	0,9998
Ложь	0,00002	0,9998

Таблица 6

Критерии X_1 (повторное посещение) и X_5 (IP-адрес изменился)

X_1, повторное посещение	X_5, IP-адрес изменился	
	Истина	Ложь
Истина	0,03	0,97
Ложь	1	0

Таблица 7

Критерии X_2 (используется SSL/TLS) и X_9 (действительный сертификат)

X_2, используется SSL/TLS	X_9, действительный сертификат	
	Истина	Ложь
Истина	0,00002	0,9998
Ложь	0,00002	0,9998

Таблица 8

Критерии X_1 (повторное посещение), X_{15} (некоммерческий удостоверяющий центр), X_{16} (коммерческий удостоверяющий центр) и X_6 (удостоверяющий центр сменился)

X_1, повторное посещение	X_{15}, некоммерческий удостоверяющий центр	X_{16}, коммерческий удостоверяющий центр	X_6V_6, удостоверяющий центр сменился	
			Истина	Ложь
Истина	Истина	Ложь	0,005	0,995
Истина	Истина	Истина	0	1
Истина	Ложь	Ложь	0	1
Истина	Ложь	Истина	0,002	0,998
Ложь	Истина	Ложь	1	0
Ложь	Истина	Истина	0	1
Ложь	Ложь	Ложь	0	1
Ложь	Ложь	Истина	1	0

Таблица 9

Критерии X_6 (повторное посещение), X_2 (используется SSL/TLS) и X_7 (сертификат сменился)

X_6, повторное посещение	X_2, используется SSL/TLS	X_7, сертификат сменился	
		Истина	Ложь
Ложь	Ложь	0	1
Ложь	Истина	1	0
Истина	Ложь	0	1
Истина	Истина	0,005	0,995

Таблица 10

Критерии X_1 (повторное посещение), X_2 (используется SSL/TLS) и X_8 (версия SSL/TLS изменилась)

X_1, повторное посещение	X_2, используется SSL/TLS	X_8, версия SSL/TLS изменилась	
		Истина	Ложь
Ложь	Ложь	0	1
Ложь	Истина	1	0
Истина	Ложь	0	1
Истина	Истина	0,003	0,997

Таблица 11

Критерии X_2 (используется SSL/TLS) и X_{10} (наличие HSTS)

X_2, используется SSL/TLS	X_{10}, наличие HSTS	
	Истина	Ложь
Истина	0,03	0,97
Ложь	0,00002	0,9998

Таблица 12

Критерии X_5 (IP изменился) и X_{11} (IP-адрес другого региона)

X_5, IP изменился	X_{11}, IP-адрес другого региона	
	Истина	Ложь
Истина	0,32	0,68
Ложь	0,00001	0,9999

Таблица 13

Критерии X_6 (удостоверяющий центр изменился)
и X_{12} (удостоверяющий центр сменился на зарубежный)

X_6, удостоверяющий центр изменился	X_{12}, удостоверяющий центр сменился на зарубежный	
	Истина	Ложь
Истина	0,00003	0,9997
Ложь	0	1

Таблица 14

Критерии X_7 (сертификат сменился)
и X_{13} (предыдущий сертификат действителен)

X_7, сертификат сменился	X_{13}, предыдущий сертификат действителен	
	Истина	Ложь
Истина	0,004	0,996
Ложь	0,00001	0,9999

Таблица 15

Критерии X_9 (действительный сертификат)
и X_{14} (самоподписанный сертификат)

X_9, действительный сертификат	X_{14}, самоподписанный сертификат	
	Истина	Ложь
Истина	0,003	0,997
Ложь	0,00001	0,9999

Таблица 16

Критерии X_9 (действительный сертификат)
и X_{15} (некоммерческий удостоверяющий центр сертификации)

X_9, действительный сертификат	X_{15}, некоммерческий удостоверяющий центр	
	Истина	Ложь
Истина	0,04	0,96
Ложь	0,00001	0,9999

Таблица 17

Критерии X_9 (действительный сертификат)
и X_{16} (коммерческий удостоверяющий центр сертификации)

X_9, действительный сертификат	X_{16}, коммерческий удостоверяющий центр	
	Истина	Ложь
Истина	0,96	0,04
Ложь	0,00001	0,9999

Таблица 18

Критерии X_{15} (некоммерческий удостоверяющий центр),
X_{16} (коммерческий удостоверяющий центр сертификации)
и X_{17} (наличие сети взаимодоверяющих удостоверяющих центров сертификации)

X_{15}, некоммерческий удостоверяющий центр	X_{16}, коммерческий удостоверяющий центр сертификации	X_{17}, наличие сети взаимодоверяющих удостоверяющих центров	
		Истина	Ложь
Ложь	Ложь	0,0001	0,9999
Ложь	Истина	0,6	0,94
Истина	Ложь	0,2	0,98
Истина	Истина	0,0001	0,9999

Одно из решений проблемы неточности значений вероятностей – это их уточнение во время использования байесовской сети. Аккумулируя пользовательские статистические данные, можно добиться высокой точности, а при наличии достаточного количества пользователей по всему миру можно решить и проблему территориальной распределенности.

Математическая модель обнаружения и предотвращения фишинговых атак перед аутентификацией пользователей, используя байесовскую сеть, может служить источником данных для традиционных черных списков и репутационных систем [4, 5].

$$P(x,z\,|\,y)=\frac{P(x,y,z)}{P(y)}=\frac{P(x)P(y\,|\,x)P(z\,|\,y)}{P(y)}=P(x\,|\,y)P(z\,|\,y).$$

Имея графическую модель и таблицы условных вероятностей, можно ответить на вопрос, с какой вероятностью Web-ресурс является безопасным для пользователя, который собирается ввести аутентификационные данные для пользования контентом, осуществлять передачу и обработку персональных данных, данных для ведения банковских транзакций [5]:

$$P(X_0\,|\,\{X_i=истина\})=P(\{X_i=истина\}),\ X_0=истина/\,P(\{X_i=истина\}).$$

Приведем пример расчета вероятности по основным 5-ти родительским показателям дерева, без декомпозиции предков и их замыкания:

$$P(\overline{V}_0.V_1V_2V_3\overline{V}_4)=P(V_0)*P(V_1\,|\,V_0)*P(V_2\,|\,V_0)*P(V_3\,|\,V_0)*P(V_4\,|\,V_0)=$$

$$=0{,}9997*0{,}9998*0{,}76**0{,}97*0{,}9998\approx 0{,}74\,.$$

В зависимости от вычисленной величины функции вероятности P результатом выполнения алгоритма может стать одно из перечисленных ниже значений:

– положительное, в случае $0\le P\le 0{,}4$ – отображается «зеленый» индикатор, сообщающий о положительном результате проверки, новые значения свойств запоминаются;

– нейтральное, если $0{,}4<P<0{,}6$ – выводится предупреждение о возможной сетевой атаке, анализе трафика; работа Web-приложения не блокируется, но пользователь получает уведомление;

– негативное, если $0<P\le 1$ – выводится оповещение о сетевой угрозе, работа с Web-приложением прекращается.

Применение данного алгоритма реализовано в плагине для Web-браузера “Google Chrome” (рис. 2). Однако в целях оптимизации вычислительных мощностей рекомендуется использовать данный плагин при открытии Web-страниц, содержащих авторизационные формы и поля заполнения персональных данных.

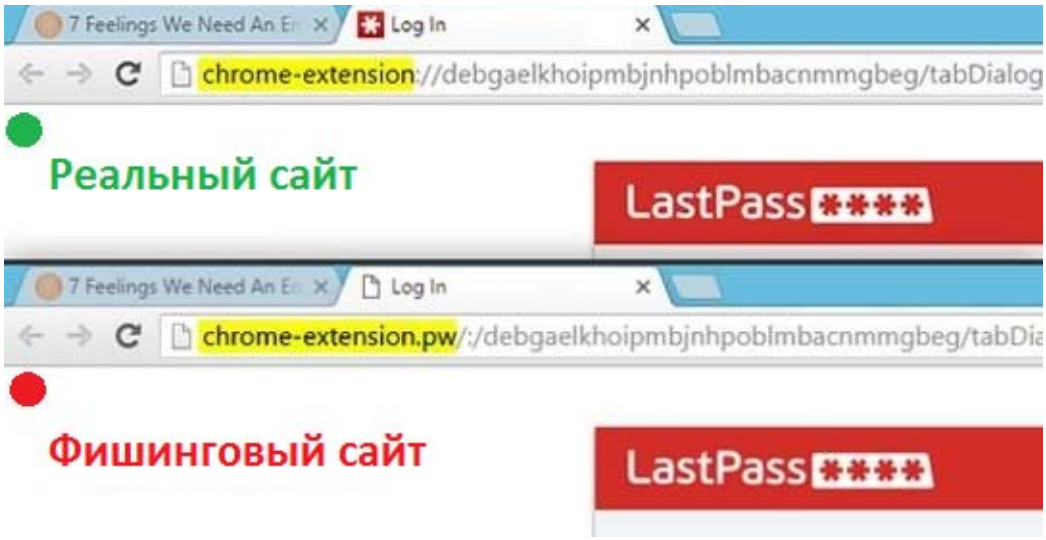

Рис. 2. Пример работы разработанного плагина-расширения в “Google Chrome” перед вводом аутентификационных данных

Примечания:

1. Власенко А.В., Дзьобан П.И. Разработка и системный анализ математической модели угроз, модели нарушителя, процедур защиты Web-приложений на всех этапах функционирования // Политематический сетевой электронный научный журнал КубГАУ = Scientific Journal of KubSAU. 2014. № 101. С. 2154–2164.
2. О персональных данных: Федеральный закон от 27.07.2006 г. № 152-ФЗ (ред. от 21.07.2014) (с изм. и доп., вступ. в силу с 01.09.2015) // СПС КонсультантПлюс. М., 2017.
3. Об утверждении требований к защите персональных данных при их обработке в информационных системах персональных данных: постановление Правительства РФ от 01.11.2012 г. № 1119 // СПС КонсультантПлюс. М., 2017.
4. Власенко А.В., Чебанов А.С., Жук Р.В. Методический подход к выбору и разработке моделей оценки эффективности комплексной системы объектов защиты // Известия Юго-Западного государственного университета. 2012. № 2. С. 038–040.
5. Власенко А.В., Дзьобан П.И. Разработка алгоритмов и программ выбора оптимального набора компонент нейтрализации актуальных угроз на основе описания модели и интеграции их в Web-приложение // Вестник Адыгейского государственного университета. Сер. Естественно-математические и технические науки. 2014. Вып. 3 (142). С. 189–193. URL: http://vestnik.adygnet.ru

References:

1. Vlasenko A.V., Dzoban P.I. Development and system analysis of the mathematical model of threats, the model of the intruder, procedures for protecting Web-applications at all stages of functioning // Polythematic Network Electronic Scientific Journal of KubSAU. 2014. No. 101. P. 2154–2164.
2. On Personal Data: Federal Law of July 27, 2006. No. 152-FZ (amended on July 21, 2014) (amended and supplemented, effective from September 1, 2015) // SPS ConsultantPlus. M., 2017.
3. On the approval of the requirements for the protection of personal data when processing them in information systems of personal data: RF Government Decree of 11.11.2012. No. 1119 // SPS ConsultantPlus. M., 2017.
4. Vlasenko A.V., Chebanov A.S., Zhuk R.V. Methodological approach to the selection and development of models for assessing the effectiveness of a complex system of objects of protection // News of the South-West University. 2012. No. 2. P. 038–040.
5. Vlasenko A.V., Dzoban P.I. Development of algorithms and programs to choose the optimal set of components of actual threat neutralization on the basis of model description and their integration in Web appendix // The Bulletin of the Adyghe State University. Ser. Natural-Mathematical and Technical Sciences. 2014. Iss. 3 (142). P. 189–193. URL: http://vestnik.adygnet.ru