On Privacy Protection in the Internet Surveillance Era

Dijana Vukovic[1,2], Danilo Gligoroski[1] and Zoran Djuric[2]
[1]*Department of Telematics, NTNU, O.S. Bragstads plass 2B, Trondheim, Norway*
[2]*Department of Computer Science and Informatics, Faculty of Electrical Engineering, Patre 5, Banja Luka, Bosnia and Herzegovina*

Keywords: Internet Surveillance, Diffie-Hellman Key Exchange, Privacy, Chat Communication.

Abstract: Snowden's whistleblower from the last year made people more aware of the fact that we are living in the Internet surveillance era. Privacy of Internet communication has been disrupted. In this paper, application for privacy protection in chat communication, named CryptoCloak, is presented. CryptoCloak provides privacy protection for chat communication. Encrypted communication is masked with dynamic cheap chat conversation. Communication made this way is not point of interest for mass surveillance spying engines. For implementation of the CryptoCloak, Facebook Messenger API is used. Diffie-Hellman key exchange is done in clandestine manner - instead of sending uniform sequence of numbers, sentences are sent. Current version provides encryption/decryption mechanism for the chat communication using strong symmetric algorithm AES in CBC mode. 256 bits of Diffie-Hellman exchanged key are used for AES-CBC.

1 INTRODUCTION

The whistleblower by Edward Snowden about Internet surveillance in the past year had huge influence on everyone: ordinary users, researchers, companies. Ordinary users started to care more about their privacy on the Internet (Hattem, 2014) researchers in the field of information security focused their research on protection against Internet surveillance, while companies started to care more about privacy and security of their users (Ashford, 2014). *The Guardian* published and keeps on publishing many articles related to this topic. The whole timeline of information discovered about NSA (US National Security Agency) surveillance can be found here (Gidda, 2013). Similar violation of users' privacy was done by GCHQ (UK Government Communications Headquarters) Information about PRISM and Tempora programs used by US and UK intelligence agencies to collect private data were the part of Snowden's whistleblower. PRISM is a tool used by the NSA to collect private electronic data belonging to the users of the major Internet companies (Google, Facebook, etc). Tempora program provides GCHQ a direct access to large amounts of the global Internet data over UK-based fibre optic cables which include transatlantic cables that carry Internet traffic between the US and Europe (Harding, 2014). According to (Khudayer et.al., 2014) NSA collects content, metadata, and upstream data, and shares it with GCHQ, Central Intelligence Agency, and Department of Justice FBI. Data are collected from different sources (Apple, Google, Microsoft, etc.) and collection is justified by Patriot Act (Section 215) and FISA Amendments Act (Section 702).

Surveillance can be defined as "close observation of a person or group, especially one under suspicion" (Martin, 2010). Many terroristic activities during past years lead to Internet surveillance. Law enforcement agencies needed the ability to conduct electronic surveillance to prevent crime, terrorism, or any kind of malicious activities exploiting the Internet. Many people have opposed surveillance because it can be considered as an invasion of privacy (as with hidden video cameras) or a tool of social control (as in monitoring workers). Surveillance can be justified in some cases, as a cracking down a crime, or increasing efficiency of service systems, but it can also be a big threat to privacy. Privacy can simply be defined as "the right to be left alone". Privacy is the right of each individual, and it should not be threatened if it is not harmful to the others.

On February 11, 2014 "The Day We Fight Back" was organized by EFF (Electronic Frontier Foundation) against mass surveillance (Kamdar,

Vukovic D., Gligoroski D. and Djuric Z..
On Privacy Protection in the Internet Surveillance Era.
DOI: 10.5220/0005043002610266
In *Proceedings of the 11th International Conference on Security and Cryptography* (SECRYPT-2014), pages 261-266
ISBN: 978-989-758-045-1

2014). Over 850,000 people took part in it and different events were held on five continents - from Ireland to India, California to South Africa. A campaign calling for a “free, open and truly global Internet” was launched by Tim Berners-Lee to mark the 25th anniversary of World Wide Web invention. It becomes evident that Internet surveillance era has brought a huge increase of the violation of privacy.

As a response to the violation of privacy, Bruce Schneier wrote the article "*The US government has betrayed the Internet. We need to take it back*." (Schneier, 2013). The main idea of this article was that the engineering community need to bring back Internet as it used to be to the people - (Schneier, 2013) "*This is not the Internet the world needs, or the Internet its creators envisioned. We need to take it back. And by we, I mean the engineering community. Yes, this is primarily a political problem, a policy matter that requires political intervention. But this is also an engineering problem, and there are several things engineers can – and should – do.*" Guided by this idea, we started the *CryptoCloak* project.

The basic idea of the CryptoCloak can be described as the following: we will use the solid and secure algorithms that have been proved as secure, but do the encryption in a clandestine manner. The automatic filtering rules of the spying agencies will notice just a cheap chat conversations, while the real encrypted information will be incorporated deeply in that cheap chat. The cheap chat will be our cloak to encrypted information.

According to (Regalado, 2013), computer scientists are involved in enabling intrusion on individual privacy and this lead to breaking the code of ethics (ACM Council, 1992). This project will be our attempt to follow the basic idea of (Schneier, 2013) and to enable use of the code of ethics in the right way – to contribute to society and human well-being without doing harm to others, and with respecting the privacy of others.

Section 2 gives an overview of existing solution related to the privacy protection in Internet communication. In Section 3 the idea and implementation of the CryptoCloak project are described, with an example of its usage. Section 4 gives an overview of further work, and the paper is concluded in Section 5.

2 RELATED WORK

To the best of our knowledge, there are no chat applications that provide private communication the way the CryptoCloak does. Most of related solutions use steganography - the act of concealing data in plain sight. Steganography application in network provides a possibility to carry on hidden information over Internet seemingly like innocent Internet traffic. Authors in (Mazurczyk et.al., 2013) presented results of their research in the field of network steganography to show how the network steganography can be used exploiting a common use of the Internet. SkyDe program exploits silence packets sent during the voice communication over Skype to send hidden information. StegTorrent program exploits the weakness of BitTorrent, that BitTorrent user often shares a data file (or pieces of the file) with many recipients at once. The third presented research resulted with StegSuggest steganography program targets the feature Google Suggest, which lists the 10 most popular search phrases given a string of letters the user has entered in Google’s search box. Wireless Padding, or WiPad method is related to Wi-Fi Networks weaknesses, on networks that use the data-encoding technique known as OFDM (orthogonal frequency-division multiplexing).

Internet censorship by government becomes an increasingly common practice worldwide. Between Internet users and censors now the "arms race" is started. For encrypted conversation over the Internet a lot of applications can be found. The most known one is Tor (Tor, 2014). It can be described as a "network of virtual tunnels". Tor provides protection from a common form of Internet surveillance known as "traffic analysis" by distributing transactions over several places on the Internet. This idea is similar to twisty - hard to follow route in order to throw of somebody who is tailing you. As a camouflage proxy for Tor, StegoTorus was developed. StegoTorus improves the resilience of Tor to fingerprinting attacks and delivers usable performance (Weinberg et.al., 2012).

Cryptocat (open source software) (CryptoCat, 2014) uses modern web technologies to provide easy to use, accessible encrypted chat. It is developed as plug-in for most popular web browsers. Chat conversation is encrypted before sending — even the Cryptocat network itself can't read it.

Spying engines do the traffic analysis in the following way: (1) filter the content looking for particular keywords, e.g. bomb, terrorism, etc.; content will be analyzed to prevent potential terroristic attacks or similar issues, (2) any encrypted content will be stored for further analysis, (3) noticing cheap chat conversations, e.g. "Hello!", "How are you?" - will be ignored. To provide spying

engines proof communication over the Internet, it has to be similar to the cheap chat communication. That can be used as a manner of masking previously encrypted communication.

CryptoCloak does not use the steganography. It does not embed the information in some other existing information. It produces a fake real-time, dynamic cheap chat and there it embeds the secret information. Messages sent via CryptoCloak application are not encrypted, as it is the case for Cryptocat, and they will not be detected by spying engines as suspicious.

3 THE CryptoCloak PROJECT

CryptoCloak communication between Alice and Bob is represented in Figure 1. Diffie-Hellman key exchange is done in the modified way - instead of regular primes the sequence of sentences is going to be sent. The first step is selecting the source of sentences: URL (Uniform Resource Locator) or database. URL has to be a link for the free online available text document to provide the same source of sentences for both communication sides (Alice and Bob) for encryption and decryption. Currently, the CryptoCloak uses the free e-books published under the "Project Guttenberg" (Gutenberg, 2014).

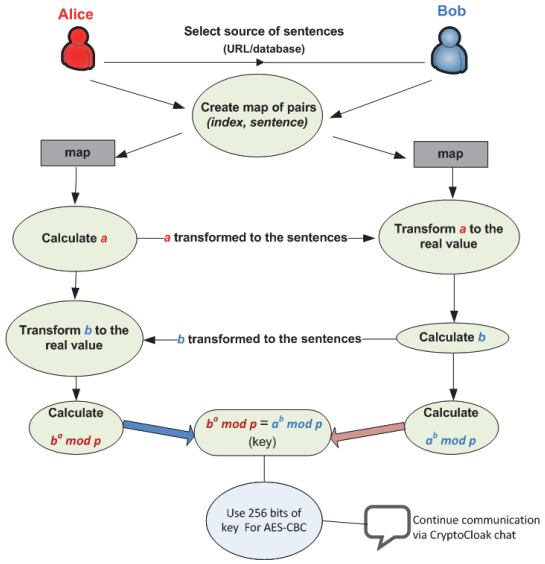


Figure 1: The CryptoCloak.

There can be found over 40000 e-books from different genre. When URL is selected as a source of sentences, Alice chooses one e-book in textual format from the "Project Guttenberg" official web site and copies its URL , and sends this URL to Bob, e.g. "Alice's Adventures in Wonderland" - http://www.gutenberg.org/cache/epub/28885/pg28885.txt. On the both sides, using the same algorithm, the chosen text document is transformed into the map of pairs (index, sentence). Only different sentences will come into consideration when transformation algorithm is applied. Next step is calculation of number n that fulfil the condition $2^n \leq sn \leq 2^{n+1}$, where sn is number of sentences in the set after transformation. For "Alice's Adventures in Wonderland" sn=962, which gives us n=9.

Using Diffie-Hellman key exchange algorithm, a should be sent over the network from Alice's side. p and g are public, and they are built in the application. After calculation of a, a is transformed into its binary value, then split into blocks of n bits. For every block, decimal value is recalculated - it is an index to get corresponding sentence from generated map. Finally, instead of regular primes, sequence of sentences is sent over the network. The length of the parameter a is also sent as a sentence - sentence is chosen from the map where index value is equal to the length value. On the Bob's side reverse algorithm is executed on the received sequence of sentences and the actual values of sent primes are shown to Bob. Bob calculates b, encrypts it the same way as Alice did with a, and send it to her. He generates his key using the b value. Alice retrieves b, decrypts it and calculates the key. At the end, both sides have the same key - $b^a \bmod p = a^b \bmod p$. When the database is selected as a source of sentences, sentences will be read from the internal database. Key exchange will be done in the same way as it is previously described.

After successful Diffie-Hellman key exchange, 256 bits of the key will be used for AES-CBC encryption. The message user (Alice) enters in the CryptoCloak chat will be encrypted using AES-CBC encryption, masked and sent after that in the same way as the Diffie-Hellman parameters were in the key exchange process. On the receiver side, received sentences will be transformed to encrypted content, and after that the decryption will be done. The receiver (Bob) sees only the decrypted content.

3.1 Implementation of the Cryptocloak

CryptoCloak is Java GUI (Graphical User Interface) chat application implemented using Java Swing API (Application Programming Interface). Implementation using Java programming language provides hardware independence, as well as operating system independence.

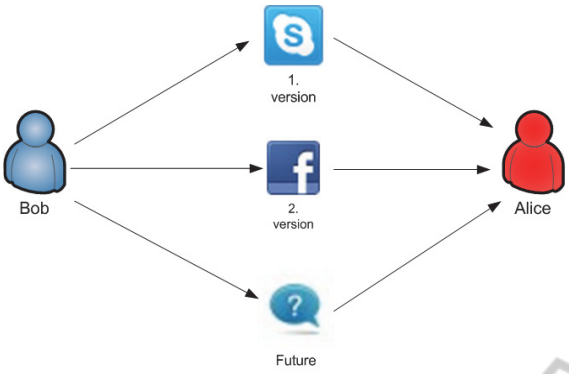


Figure 2: The versions of the CryptoCloak.

In the first version of the CryptoCloak (Figure 2), all communication between Alice and Bob was done over the Skype. In that purpose, Skype API was used. At the end of 2013, Skype API got retired (Callaham, 2013). In the second version of the CryptoCloak, Facebook Messenger API is used.

The internal structure of the CryptoCloak chat application is shown in Figure 3.

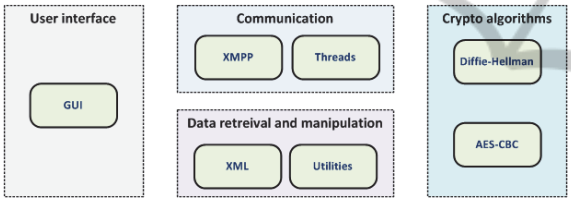


Figure 3: The CryptoCloak inner structure.

It consists of four modules:

- User interface – consists of classes that represents GUI of the CryptoCloak application,
- Communication – consists of classes for sending/receiving chat messages,
- Data retrieval and manipulation – consists of classes for preparing appropriate set of sentences,
- Crypto algorithms – consists of implementations for used cryptographic algorithms.

To establish communication over the Internet for instant messaging, XMPP (Extensible Messaging and Presence Protocol) protocol is used. XMPP is a protocol for streaming XML elements in order to exchange messages and presence information in close to real time. Additional Java library is used for XMPP communication – Smack (Jive Software, 2014). Smack is an Open Source XMPP (Jabber) client library for instant messaging and presence. The entire XMPP communication is done over the Facebook XMPP server (chat.facebook.com), and to use the CryptoCloak a legitimate Facebook account is needed. Module for communication, besides XMPP part, has a part for sending and receiving messages which lays on concurrency (Threads) – two different threads for sending and receiving messages are implemented (two per user), and it supports multithreading Research of the average typing speed from (Fort, 2014) is used as a reference for making delay between sending sentences, to make the impression that communication is not automated. According to (Fort, 2014) average typing speed of regular user is 41 words per minute. Before sending sentence, number of words in it is determined, and the approximated time for typing the sentence is calculated. After receiving sentence, some random sentence will be picked up from sentences map and sent to the other side in the same way.

As it is already described in Section 3, CryptoCloak can use two sources of sentences for cheap chat communication: online textual file, or internal database. To eliminate dependency of RDBMSs (Remote DataBase Management Server) and avoid need of additional installation and configuration on the user computer, a simple XML file is used as data store. Currently, XML database is filled with dialogues from English learning book "Speak English like an American", and from online school for practicing English conversation (http://www.eslfast.com/robot/topics/smalltalk/smalltalk.htm), and it contains 1438 sentences. Speed of conversation is dependent on the size of the map. The bigger map, the larger will be size of the block, and it will decrease the number of sentences for sending parameters a and b.

Data retrieval and manipulation module contains of XML and Utilities parts. XML part does reading and writing XML file using standard Java API. At the current version, CryptoCloak uses two crypto algorithms: Diffie-Hellman and AES-CBS. Additional cryptography library is used for the implementation of this module – Bouncy Castle (Bouncy Castle, 2014). A part for the Diffie-Hellman calculations uses SecureRandom Java class for generating needed random numbers, because commonly used Random class is not recommended for use in cryptographic implementations. Parameters p and g for Diffie-Hellman can be read/changed from/in the properties file built in with application. *The important fact is that both sides, Bob and Alice, have to agree on the source of sentences, and p-g parameters (if they are entered*

manually, they have to be the same), otherwise Diffie-Hellman key exchange will not be successful.

In the Utilities part different Java classes with methods for reading properties file, parsing textual file to get different sentences, conversing primes to sentences array etc., can be found.

3.2 Usage of the CryptoCloak

CryptoCloak is a simple GUI application. After running application, login window appears (Figure 4). To start using CryptoCloak chat, login has to be done with Facebook login credentials – username and password.

Figure 4: The CryptoCloak login window.

For testing purpose, two new Facebook accounts were created. In the testing phase, application was installed on one PC with Intel(R) Core(TM) i7-3632QM CPU @ 2.20GHz with 8 GB of RAM, and on the Virtual machine on that PC. Host operating system was Windows 7 Home Premium with Service Pack 1, and virtual operating system was Ubuntu 12.04. 1GB of host RAM was allocated for virtual operating system.

When the parameter *p*, needed for Diffie-Hellman key exchange, is sending over the CryptoCloak chat application, it will be masked with the array of sentences as it is described in Section 3. If we have *p* set in the properties file with a value "167874266979851516265177795716240899830221206662716522903333003798255278973984252844018061182803659376670161982983927354009589606120936489674186482792573720065307230542249474500110088297085132682140410781560309403079138392437468278869927414011500536583742199605534369256553941885657744547015267866266696066601", array of 103 sentences will be sent. In Figure 5, one part of the sentences sending in shown. In Figure 6 the result of successful key exchange is shown.

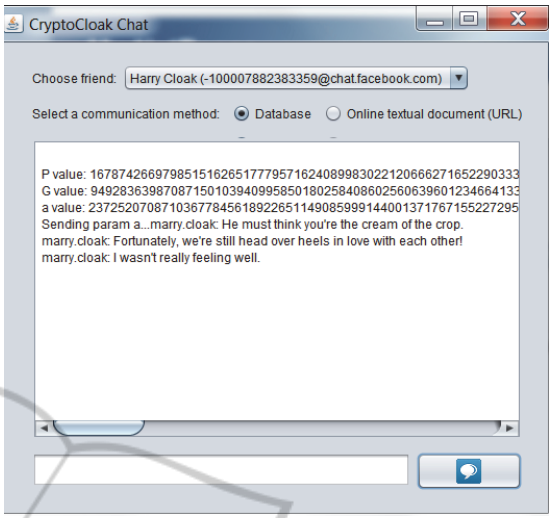
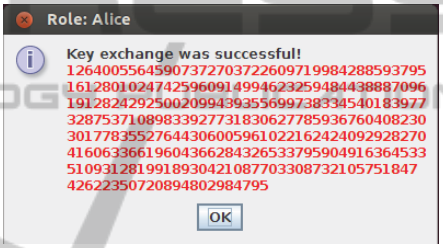
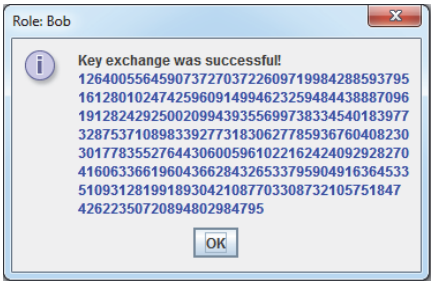


Figure 5: Sentences sending.



(a)



(b)

Figure 6: (a) Alice; (b) Bob.

4 FURTHER WORK

Exchange time for *a* and *b* parameters is quite long (in total around 30 minutes). Considering the fact that Java has an e-mail API, we can speed up a process of parameters exchange by sending them as an e-mail message. Using the same algorithm described in Section 3 parameters will be converted into array of sentences, and, instead of sending sentences via chat communication, they will be sent using legitimate e-mail account. CryptoCloak will be

modified in a way that user can send/receive e-mail message over/from different accounts. It means that, instead of sending parameters *a* and *b* over chat communication, they, as sentences array, can be split and send over a few different e-mail accounts, and merged from these accounts on receiver side. When the particular parameter is received, it will be transformed into its real value, and the key will be calculated. The same process will be executed on the both sides, Bob's and Alice's, and after the successful Diffie-Hellman key exchange, they can start AES-CBC encrypted communication in the way described in Section 3. If we split communication in a way that the key exchange will be done over e-mail and AES-CBC communication over XMPP, it will be efficient and harder to follow. It will be similar to the technique the Tor uses – twisty (route which is hard to follow).

Facebook announced shutting-down of Facebook messenger. The question mark in Figure 2 stands for the new way of communication in the future that will be used instead of current Facebook Messenger API. Considering the fact that XMPP is reliable protocol, "?" can be replaced with any free available XMPP server.

5 CONCLUSIONS

Privacy of individuals should not be threatened in any case, but Internet surveillance era represents a huge threat to it. The CryptoCloak project has the aim – protection against surveillance in the chat communication. Diffie-Hellman key exchange over the network, without sending sequence of bytes, will not be detected by traffic analysis tools (spying engines). After successful key exchange, AES is used for encryption/decryption of chat communication, and the CryptoCloak masks communication providing dynamic cheap chat conversation. The CryptoCloak is now in the testing and bugs fixing phase, and some additional features (e.g. selection of conversation topic, different languages support) will be added in the future.

ACKNOWLEDGEMENTS

Dijana Vukovic, as a PhD student in the field of information security, is supported by the COINS Research School of Computer and Information Security.

REFERENCES

Hattem J., 2014. "Many say NSA news changed their behavior", The Hill.

Ashford W., 2014, "Yahoo encrypts users' data to boost security and privacy after NSA revelations", The ComputerWeekly.

Gidda M., 2013. "Edward Snowden and the NSA files – timeline", The Guardian.

Harding L., 2014, "The Snowden Files: The Inside Story of the World's Most Wanted Man", Guardian Books & Faber and Faber, UK, first edition.

Kamdar A., 2014. "Today We Fight Back Against Mass Surveillance", EFF.

Schneier B., 2013. "The US government has betrayed the internet. We need to take it back.", The Guardian.

Regalado A., 2013. "Cryptographers Have an Ethics Problem", MIT Technology Review.

ACM Council, 1992. "Code of Ethics", ACM.

Martin B., 2010. "Opposing Surveillance", IEEE Technology and Society Magazine, 29 (2), pp. 26-32.

Mazurczyk W., Szczypiorski K., Lubacz J., 2013. "4 New Ways to Smuggle Messages Across the Internet", IEEE Spectrum.

Weinberg Z., et.al, 2012. "StegoTorus: a camouflage proxy for the Tor anonymity system", in Proceedings of ACM CCS '12, pp. 109-120.

Callaham J., 2013. "Skype to retire Desktop API support by end of 2013", Neowin.

CryptoCat, 2014. https://crypto.cat/.

Jive Software, 2014, http://www.igniterealtime.org/projects/smack/.

Gutenberg, 2014. http://www.gutenberg.org/.

Khudayer A., Abdulsalam R., Alshaibani S., Bin Ibrahim J., 2014. "Impact of NSA-PRISM to National Information Security Strategy & Policy", International Journal of Information and Communication Technology Research, Volume 4 No. 1, January 2014, ICT Journal.

Bouncy Castle, 2014. https://www.bouncycastle.org/.

Fort A., 2014. "Why Average Typing Speed is Important?", eLearning Industry.