New generation data- and network protection procedures

Theses of the PhD dissertation

Péter Vörös

Supervisor: Attila Kiss, Ph.D., C.Sc.

Eötvös Loránd Uniersity
Faculty of Informatics
Department of Information Systems

Ph.D. School of Computer Science
Head of school: Prof. Erzsébet Csuhaj-Varjú
Ph.D. Program of Information Systems
Head of Program: Prof. András Benczúr

Budapest, 2019

Introduction

Due to the spread of smartphones and various online services, people today are generating an incredible amount of data. According to Cisco's new Visual Networking Index (VNI), Internet traffic will be greater in 2022 than in the previous 32 years since the launch of the Internet. With ever-smarter gadgets, people are sharing more and more sensitive data, intentionally or unwittingly. The huge datasets, and services that are responsible for managing them face experts with new data storage and data security challenges.

The current technology complexity is very complicated, therefore for the general users it is hard to be able to know the details about how these systems operate. On one hand, this is quite understandable, because everyone cannot be be an IT expert, but at the same time, it is very unfortunate that the majority of the average Internet users have no idea what is considered to be safe, or what the risks of giving one's credit card data on a phishing site. Luckily, another group of users is starting to grow up, who are seriously concerned about the need to keep their data safe.

In this dissertation, I present the most common types of attacks against current services, the structure of attacks, and I show various protection models for both service providers and users. I identify the different security points, and I also present the tools/services that I have designed at these points in order to preserve the safety of sensitive data.

Theses

1 Safety methods for service providers

New generation packet forwarding in network security

The key to build long-term systems is to be easily expandable and configurable for tomorrow's needs. Currently, we are moving towards an era where every device is smart and everything is connected with everything. That's why we need to make our network systems flexible and easily scalable. The classical vertically integrated networks had not made it possible to change existing protocols without having to do an incredible amount of extra work.

These traditional, inflexible networking systems for fixed protocols have been gradually replaced by SDN [15][14]. (Software Defined Networking), granting network operators a new level of programmatic control over their networks.

Amongst different high-level languages, P4 [12] (Programming Protocol-independent Packet Processors) has been proposed to overcome the limitation of fixed protocols by allowing the developers to describe the packet forwarding in an abstract, protocol independent way. Compared to OpenFlow, it provides a much higher level of abstraction for network programming.

Thesis 1. P4 makes it is possible to implement efficient protocol- and hardware-independent firewalls with a high abstraction level.

Related papers: [3] [4] [6] [7] [9]

In Chapter 1.1 of the dissertation, I presented the opportunities offered by the next generation packet forwarding, and the P4 language which is a high abstraction domain specific language for describing packet forwarding logic. I have explained the operation of our P4 compiler named T4P4S, I have shown that the compiled switch program is capable to forward packets at nearly the same rate as hardware optimized binaries. Then I have shown a possible method to use P4, as a second-generation - packet filter firewall.

Modeling of DDoS attacks and IDSs

Creating effective defense strategies in IT security is a must, also it is often a great challenge. According to the 2014 Cyberthreat Defense Report, [18], which involved more than 750 security decision-makers and professionals, more than 60% of organizations were attacked in 2013.

The big data analysis in security gives us the possibility to collect and analyze the massive amount of digital data and to be able to predict and to prevent attacks. However, the collection of the required data in an efficient and reliable way is not trivial. Therefore, a tool that can measure and improve the effectiveness of security algorithms may be useful for the industry. For this purpose, I present a platform which is capable of generating a greatly configurable simulated Internet traffic, with a combination of attack-free and malicious network traffic patterns. For this work, I've used the ns3 event-driven network simulator. In order to meet the benchmarking objectives of the intrusion detection system, we examined the statistical characteristics of normal and malicious traffic patterns.

Thesis 2. It is possible to design a data generator for IDSs validation that can generate HTTP traffic with specific server characteristics.

Related papers: [2]

In Chapter 1.2, I modelled HTTP flows and showed how DDoS attacks are built up. I've designed a procedure to model the flood-type DDoS attacks. By preserving the characteristics of HTTP traffic, I have created a richly parameterizable traffic generator that can be used to test Intrusion Detection Systems (IDS).

2 User privacy in public clouds

Server-side protection

As a service provider, one of the most important tasks is to keep users and their data safe, but practice shows that websites are often vulnerable to the most known attack types. A large number of websites only use encrypted channel (HTTPS) on the login page, which is understandable because the encryption of the entire communication may be too resource intensive, but also the lack of encryption results in a serious vulnerability. Especially if the provider does not protect session cookies enough.

HTTP is a stateless protocol, which means users can only be recognized from a specifically issued user ID. This is stored on the client side as a cookie. The specific cookie that identifies the user session is called a session cookie. Session-hijacking is a type of attack where an attacker somehow obtains the victim's session cookie. Then this cookie will allow the attacker to fool the server and be identified as the victim. So, he gets access to do everything using the victim's rights.

Thesis 3. A One Time Token-based server-side security framework can significantly improve protection against Session hijacking attacks.

Related papers: [5]

In Chapter 2.1, I've introduced the session-hijacking and data-breach attacks, which are 2 of the 12 highest-priority threats. I analyzed the dangers of HTTP and HTTPS traffic, some possible methods of the hijacking attacks, and outlined a new one-time token based authentication model. I have shown that my software called TooKie with single-use tokens secures user sessions even over HTTP.

Client-side and proxy-based protection

The amount of user-produced data is increasing day by day. Social networks, IoT, sensors, etc. are all responsible for this growth. Unfortunately, many users are not aware of the importance of their own data. Companies spend a significant amount of money on securing their data, according to their business sector, size, and budget, because they need to protect their industrial secrets, their intellectual property, and so on. We believe that the same applies to users, due to the fact that a large proportion of these attacks target their data. Since not only company data represents business value, but also average users'. Forrester Consulting, in their study, presented the most valuable customer identification data for marketing purposes [17]. CSO reports [16] that billions of people's data were stolen by 2018, which also means that their data needs to be secured.

In terms of data storage trends, we found that in addition to users, some companies also use third-party storage, most of which are cloud-based. The 2014 EU statistics show that on average 21 percent of the people in the region use cloud storage to store files, while the maximum value is 42 percent [13]. For end users, Dropbox, OneDrive, and Google Drive are among the most popular choices due to their high availability and universal accessibility. And of course it is also a crucial that these providers offer free storage capacity. Despite its data security measures, the user cannot really influence how they store and protect their data, so the service provider may have access to the user's data. A third-party access channel may also be compromised if the provider does not guarantee security. If this happens, the data can become accessible to the man-in-the-middle attackers.

There are a number of suggestions and solutions for keeping our data safe, but unfortunately, nothing has been widely accepted so far. I believe that the main reasons are the complexity of the software and the lack of knowledge and awareness of the users in the security area. In this study, my goal was to provide an easy-to-use client-side method to make user data inaccessible to 3rd persons in public clouds.

Thesis 4. The OpenWebCrypt browser extension application and the CrypStorePI security middleware are suitable for protecting individual user data or the entire private network against Data Breach attacks.

Related papers: [1] [8] [10]

In the second part of Chapter 2, I've presented methods for implementing an extra layer of protection for general users and larger corporate networks. The OpenWebCrypt

client-side browser extension can encode user data stored in various cloud services. Also, the complete network can be protected by CrypStorePI which is a proxy-based security middleware.

Steganography and cryptography for user data in calendars

Users generate large amounts of personal data daily on online services such as social networks, search engines, and more. without asking themselves how valuable is my data to the service providers? As the collection and classification of personal data is an essential part of targeted advertising, the answer is very valuable. Marketers increasingly rely on customer data, so their value has increased significantly over the past decade. More and more data is stored in different clouds and the users have no tool to guarantee their safeness, as most of the service providers do not offer any way to set up reliable encryption. In this study, I present a possible way to hide personal information from curious eyes.

We have created a two-layer security model to protect one's personal information. The first layer is steganography. Steganography is the practice of concealing a file, message, image, or video within another file, message, image, or video. We can also use this to deceive the attacker and conceal the very existence of the encoding at all. Katzenbeisser et. al present [11] steganography in details. We use a dictionary-based algorithm that generates encoded text data that looks like normal user input. It is perfect to mislead the attackers and hide sensitive data. The second layer is a simple, personal password-based encryption that encodes the data with a simple stream cipher.

Thesis 5. It is possible to design a process that encrypts calendar entries to other seemingly meaningful entries, concealing the existence of the encryption.

Related papers: [10]

At the end of Chapter 2, I have shown a practical example of how steganography can be used for certain services to conceal the encryption of data from the attackers.

Related publications of the author

[1] Vörös Péter and Kiss Attila. "Felhő architektúrák biztonsága". In: *INFODIDACT 2014: Informatika Szakmódszertani Konferencia. Konferencia*. Vol. 16. 9789631206272. Webdidaktika Alapítvány. 2014, 9–p.

[2] Dániel Csubák et al. "Big data testbed for network attack detection". In: *Acta Polytechnica Hungarica* 13.2 (2016), pp. 47–57.

[3] Sándor Laki et al. "High speed packet forwarding compiled from protocol independent data plane specifications". In: *Proceedings of the 2016 ACM SIGCOMM Conference*. ACM. 2016, pp. 629–630.

[4] Péter Vörös and Attila Kiss. "Security middleware programming using P4". In: *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, Cham. 2016, pp. 277–287.

[5] Péter Vörös and Attila Kiss. "TooKie: A New Way to Secure Sessions". In: *Recent Developments in Intelligent Information and Database Systems*. Springer, 2016, pp. 195–207.

[6] Tamás Lévai et al. "The Price for Programmability in the Software Data Plane: The Vendor Perspective". In: *IEEE Journal on Selected Areas in Communications* 36.12 (2018), pp. 2621–2630.

[7] Fabricio Rodriguez et al. "BB-gen: A packet crafter for P4 target evaluation". In: *Proceedings of the ACM SIGCOMM 2018 Conference on Posters and Demos*. ACM. 2018, pp. 111–113.

[8] Péter Vörös and Attila Kiss. "OpenWebCrypt—Securing Our Data in Public Cloud". In: *Modern Approaches for Intelligent Information and Database Systems*. Springer, 2018, pp. 479–489.

[9] Péter Vörös et al. "T4P4S: A Target-independent Compiler for Protocol-independent Packet Processors". In: *International Conference on High Performance Switching and Routing* (2018).

[10] Péter Vörös, Péter Hudoba, and Attila Kiss. "Steganography and Cryptography for User Data in Calendars". In: *Asian Conference on Intelligent Information and Database Systems*. Springer. 2019, pp. 241–252.

References

[11] Stefan Katzenbeisser and Fabien Petitcolas. *Information hiding techniques for steganography and digital watermarking*. Artech house, 2000.

[12] Pat Bosshart et al. "P4: Programming protocol-independent packet processors". In: *ACM SIGCOMM Computer Communication Review* 44.3 (2014), pp. 87–95.

[13] *EUStats: Use of Internet Cloud storages*. 2014. URL: `http://ec.europa.eu/eurostat/statistics-explained/index.php/Internet\_and\_cloud\_services\_-\_statistics\_on\_the\_use\_by\_individuals` (visited on 02/17/2018).

[14] Nuno P Lopes et al. "Checking beliefs in dynamic networks". In: *Proceedings of the 12th USENIX Symposium on Networked Systems Design and Implementation, NSDI*. Vol. 15. 2015.

[15] opennetworking.org. *SDN Definition*. 2015. URL: `https://www.opennetworking.org/sdn-resources/sdn-definition` (visited on 10/10/2015).

[16] *CSO: Biggest data breaches of the 21st century*. 2017. URL: `https://images.idgesg.net/images/article/2017/10/biggest-data-breaches-by-year-and-accounts-compromised-1-100738435-large.jpg` (visited on 02/16/2018).

[17] *Forrester Consulting: Most valuable customer data*. 2017. URL: `https://www.marketingcharts.com/wp-content/uploads/2017/06/LiveIntentForrester-Most-Valuable-B2C-Customer-Identification-Data-June2017.png` (visited on 02/10/2018).

[18] *CyberEdge - 2014 Cyberthreat Defence Report for North America and Europe, a CyberEdge report sponsored by ForeScout Technologies, Inc., 2014*.