ISSN: 2597-8993

How a thief stole your data over network

Andreas Bernhard[1)], Kurniawan Hagi[1)]

Mercubuana University
asong.soekamti@gmail.com

ARTICLE INFO

Article history:
Received May 22, 2017
Revised July 15, 2017
Accepted July 29, 2017

Keywords:
Big Data
Data Theft
Networking
Packet Sniffing
Wireshark

ABSTRACT

Until now, lot of companies that already have data that is so large and the development of data is so quickly and with the plurality of data formats. Not only in units hundreds of Mega or Giga again but have reached the realm of Tera data countless. Besides the diversity of the data format itself, today not just formatted text but can be video or image as well as the format of other data. With the trend of data continues to grow, sometimes aspect of data security was often overlooked and even ignored.

I. Introduction

As already described in Abstract above, that the data is so large (volume) data is growing so fast (velocity), as well as the diversity of data formats (variety) is the core of Big Data [1] which is currently a trend that is being studied by society in the digital age.

Try to imagine a company that has Big Data, in addition to collecting data that is a time consuming and not easier in data management, maintainance the data itself is very difficult especially requires extra effort and time.

Examples of companies engaged in data management and implementing Big Data is a Google. Every service from Google itself implement the Big Data be it Youtube, Google Drive, Google Spreadsheets, Google Mail and other Google services.

For example in Indonesia, Big Data implemented are mostly located in the area of Service Provider (SP) such as Telkomsel, Indosat, XL and etc. Service Provider itself using Big Data to record every customer transaction data from both the history calls / sms, internet usage and also to record the balance pulse, as well as the use of pulses of each customer.

Companies such as service providers to meet the information processing needs are so great is necessary to implementing Big Data.

II. Data Theft

Imagine, if so much customer data belonging to one of the providers mentioned above can be stolen by irresponsible elements. Then Big Data has been collected over the years lost due data drop by thief.

Worse, if the data obtained by the thieves used for specific purposes, such as sold to competitors or used for fraud. Suppose data is stolen and used by the thief to commit acts of fraud by using information obtained from data that has been stolen. (i.e: Rustock botnet malware on behalf of Microsoft for fake lottery) [2]

The service provider could have doubtful his integrity in terms of keeping the customer confidential data. This can have a big impact on corporate earnings could be down after the incident. As we know some time ago, where there was a case that shocked the residents in the area Prapanca Raya, South Jakarta.

DOI:

Where, there is someone who can perform remote access illegally use TeamViewer in the region's existing Videotron and accidentally open profanity and videos appear on the Videotron. This will make the service provider Videotron questionable integrity and credibility. It could have been very influential in reputation and income of the service provider Videotron itself.

This is arguably the credential data theft. Where people who do not have official access to the Videotron, can illegally access with credential data that has been obtained from the admin negligence. So the data that credential such as private attribute, even visible to the public freely in videotron. [3]

Just imagine only due to an admin negligence, the reputation of a company that has earned the trust of government can be an instant destroyed. Moreover, destruction or theft of data by a thief, could lead to a more severe impact for the company

III. Data Security

Aspects of its own data security is necessary to secure the data that has been painstakingly collected, it is important to deal with such cases that have been discussed previously.

“So how important securing the data on a company?”

Let's equate the perception, that the data is a valuable asset for the company. If the assets of the company is known by its competitors, then the existing data not valuable for the company anymore. Because the data itself has been known and studied by a competitor of the company. Therefore, the security aspects of Big Data is very important to consider.

IV. Methods of Data Theft

4.1 The kind of Data Theft methods

There are various methods that are scattered on the Internet, ranging from capture traffic exchanges communication packets (packet sniffing), or forge the DNS intended to control the data thieves, those that last capture cross-exchange of data, block data and modify it later sent back to the destination and there are still other methods of data theft.

Definition 4.1.1 The Domain Name System (DNS) is a hierarchical decentralized naming system for computers, services, or any resource connected to the Internet or a private network. (in example : www.abcdefg.com) [5]

4.1.1DNS spoofing method

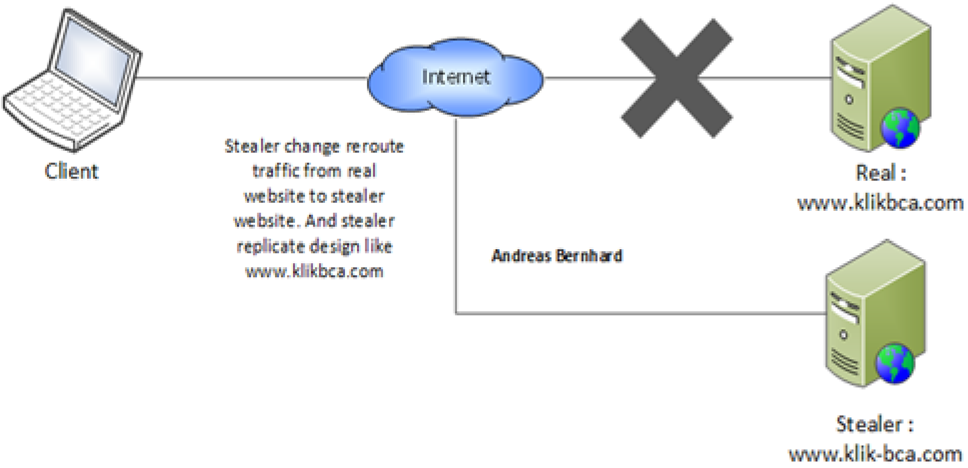

Internet
Client
Stealer change reroute traffic from real website to stealer website. And stealer replicate design like www.klikbca.com
Real : www.klikbca.com
Andreas Bernhard
Stealer : www.klik-bca.com

Fig 4.1.1

In this method, data thieves trying to make potential victims to visit the wrong site to provide confidential information to unauthorized parties in this case are the data thieves. This method works is a spoof to the original DNS server (i.e : we use www.klikbca.com for an example), when you do access to sites with www.klikbca.com such as there is nothing strange with urls accessed.

It is true that we make access to the DNS www.klikbca.com but that has been spoofed by data thieves. In other words, DNS records are made is in the data server thieves, or not www.klikbca.com official by BCA.

Nor any data and applications is not accommodated in official webserver belonging to BCA, but in webserver data thieves. And also for false contrived web applications as closely as possible to the original, it was all put on the server data thieves. Intended if the victim unknowingly already have a deal despite failing but some of the information to be obtained attacker successfully entered into the web server data thieves.

4.1.2 Packet Sniffing method

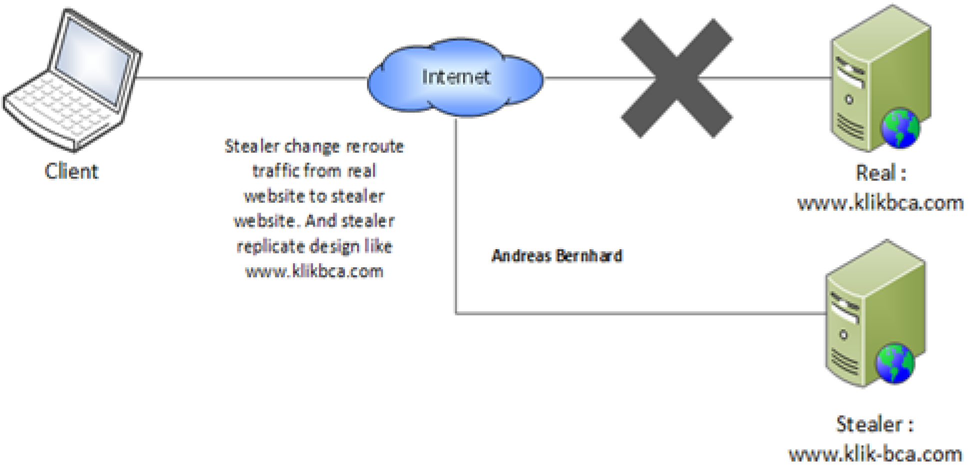

Client
Internet
Stealer change reroute traffic from real website to stealer website. And stealer replicate design like www.klikbca.com
Real : www.klikbca.com
Andreas Bernhard
Stealer : www.klik-bca.com

Fig : 4.1.2

In this method, data thieves to record the communication that occurs between potential victims with a web server on the Internet network. To overcome this problem needs to be done encryption of data packets on both sides between the client and server data. This technique is have a condition.[6]

4.1.3 Man-In-The-Middle Attack (MITM Attack) method

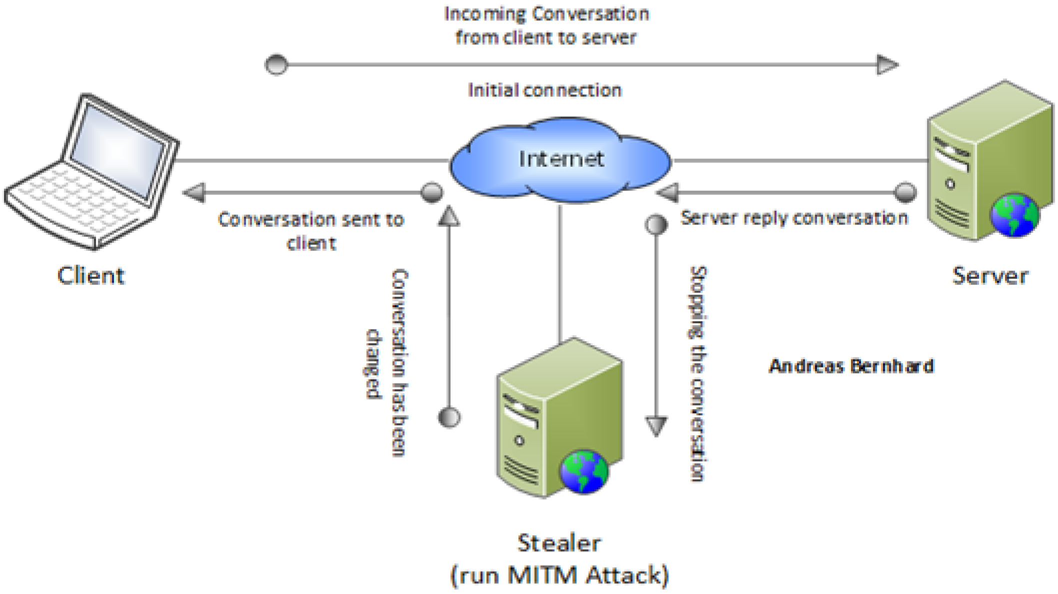

Incoming Conversation from client to server
Initial connection
Internet
Client
Conversation sent to client
Server reply conversation
Server
Conversation has been changed
Stopping the conversation
Andreas Bernhard
Stealer (run MITM Attack)

Fig. 4.1.3

If you have previously been aware of the workings of packet sniffing, in fact Man-In-The-Middle Attack The simple concept is similar to packet sniffing, namely record any communication between the client to the server and vice versa.

In contrast to packet sniffing, if likened. Packet sniffing is a passive attack and an active is MITM attack. Why are said to be active attack? Due to MITM Attack itself can block a communication received either come from the server or client and then change the contents of communication from client to server or vice versa. Once converted and then sent to the destination.

Suppose, blocking communication between server and client. Prior to the communication from the server to client, the content of such communications changed, just after it communication results that have changed are sent to the client. [7]

Of the three methods above, actually there are many methods that are on the internet to steal data. Indispensable to know the method to determine how to minimize data theft.

5. Packet Sniffing

After learning methods are often used by data thieves to perform the action. In this section only a review of how the stolen data over the network using packet sniffing.

Packet sniffing methods itself is used by practitioners of the network (network engineer) for troubleshooting the cause is difficult to solved. But over time, the method of packet sniffing changed the function of the previously only has the function for troubleshooting into an investigative tool for digital forensic or could be a , for example to study the habits of malware on the network (Each habits of malware so clearly seen on flow packet in wireshark).

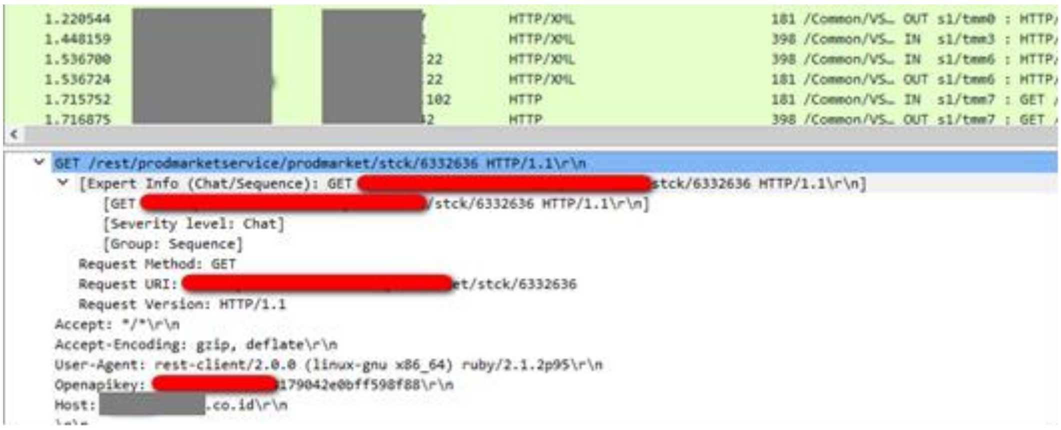
To perform packet sniffing themselves usually own data thieves to gain access, either legally or illegally into a device that his position is between the data server and the client.

Software used by data thieves to run packet sniffing method can use tcpdump, wireshark or dsniff.

Definition 5.1 Wireshark is a network packet analyzer. A network packet analyzer will try to capture network packets and tries to display that packet data as detailed as possible [8]

Terms : commit itself already sniffing method as mentioned above. That sniffing should be between the client and server data. In other words, there is a device between the two entities, then the data thief can record every communication from the client to the data server and vice versa.

After successfully recording any communication in the network, data thieves typically use wireshark to facilitate reading of the communication flow. Each communication flow so clearly seen in wireshark. In itself there is a column wireshark filter that can be used to find the communication flow with a specific format, for example, only displays the TCP communication is finished or finish in wireshark filter is tcp.flags.finish.



```
1.220544 [redacted] [redacted]       HTTP/XML    181 /Common/VS… OUT s1/tmm0 : HTTP/
1.448159                             HTTP/XML    398 /Common/VS… IN  s1/tmm3 : HTTP/
1.536700                       22    HTTP/XML    398 /Common/VS… IN  s1/tmm6 : HTTP/
1.536724                       22    HTTP/XML    181 /Common/VS… OUT s1/tmm6 : HTTP/
1.715752                       102   HTTP        181 /Common/VS… IN  s1/tmm7 : GET /
1.716875                       2     HTTP        398 /Common/VS… OUT s1/tmm7 : GET /

GET /rest/prodmarketservice/prodmarket/stck/6332636 HTTP/1.1\r\n
  [Expert Info (Chat/Sequence): GET [redacted]stck/6332636 HTTP/1.1\r\n]
      [GET [redacted]/stck/6332636 HTTP/1.1\r\n]
      [Severity level: Chat]
      [Group: Sequence]
    Request Method: GET
    Request URI: [redacted]et/stck/6332636
    Request Version: HTTP/1.1
  Accept: */*\r\n
  Accept-Encoding: gzip, deflate\r\n
  User-Agent: rest-client/2.0.0 (linux-gnu x86_64) ruby/2.1.2p95\r\n
  Openapikey: [redacted]179042e0bff598f88\r\n
  Host: [redacted].co.id\r\n
  \r\n
```

Fig. 5.1

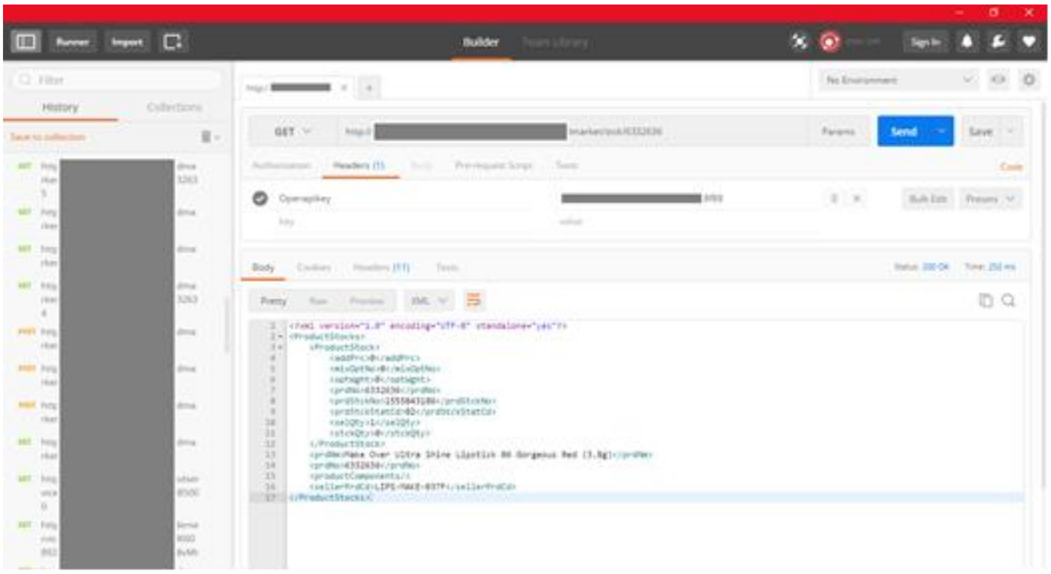
But usually focus of the data thieves is filter http.response.code with code 200. The code 200 which signifies the HTTP communication with the status OK, then the data thieves often use these filters to read the HTTP communication flow whose status is OK. For its own filter more or less as follows: http.response.code == 200 or http.response.code eq 200, Then type that filter in the filter field and press after that press Enter.

Definition 5.2 HTTP is short for HyperText Transfer Protocol. HTTP is the underlying protocol used by the World Wide Web and this protocol defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands. (i.e : http://www.abcdefg.com/)[9]

Then, the only communication flow shown just about HTTP with response code 200. After that focus of the data thieves is searching for keywords such as Request URI header, Full Request URI, Request Method, Api, OpenapiKey and others header format (according to the needs the data thieves).

So the content of the header information already mentioned above are typically noted and then inserted into an application called Postman. Then the data thieves usually combine the information obtained previously to try to get the data that him looking.

After the necessary information has been fulfilled then the information is incorporated into an application called Postman (Postman is the swiss army knife of API tools) [10]. Usually the data thieves will send an HTTP GET method REQUEST to the data server victim using Postman application to get the data that he wants.



GET http://[redacted]/market/stock/4332636

Openapikey [redacted]

Status: 200 OK Time: 250 ms

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ProductStocks>
    <ProductStock>
        <addPrc>0</addPrc>
        <mixOptNo>0</mixOptNo>
        <optWght>0</optWght>
        <prdNo>4332636</prdNo>
        <prdStckNo>1555843180</prdStckNo>
        <prdStckStatCd>02</prdStckStatCd>
        <selQty>1</selQty>
        <stckQty>0</stckQty>
    </ProductStock>
    <prdNm>Make Over Ultra Shine Lipstick 08 Gorgeous Red (3.5g)</prdNm>
    <prdNo>4332636</prdNo>
    <productComponents/>
    <sellerPrdCd>LIPS-MAKE-03TP</sellerPrdCd>
</ProductStocks>
```

Fig. 5.2

And if the requested data is valid, for example the format of data sent in accordance with the standards specified by the system or API (an application programming interface) being used is valid. The data server will reply (response) data in accordance with what is already in the REQUEST define the HTTP GET method before.

Well, cannot imagine right? if such a method can steal the data or information from your data server over the network. Imagine if something like this happened in the company that implements Big Data. Then, would carry around privacy and security of user? Therefore, why do we need to study the how a thief can steal data through the network.

6. Conclusion

Trend of data develop fast, growing big and rich with data format diversity make the BigData into an asset that is so important to the company.

Data thieves itself see the trend of Big Data as a gold mine, because this is a trend in this era. Data thieves took a chance, vying and also studied the vulnerabilities that exist in both the information system architecture, BigData system or the infrastructure of BigData itself.

Does not rule out the possibility of data theft could occur at any time, it could be because the problem is trivial but fatal, as in the case carelessness of operator Videotron who got crowded some time ago. Now therefore, we must be careful of the actions that may provoke the occurrence of data theft.

Data remains a valuable corporate assets if the data is not leaked either to the public or competitors.

Methods of data theft even for currently diverse and flourishing. One of them is the method of Man-In-The-Middle Attack (MITM Attack), which is one method that evolved from its own method of Packet Sniffing concept.

“So how to minimize the theft of data from one of the above methods?”

So how to minimize the theft of data from one of the above methods? One of them use the method that is used to anticipate packet sniffing is trying to set up the system with end-to-end encryption, so that when the data thieves managed to get any information that he can see on wireshark, extra effort is still needed to decrypt the information itself.

Acknowledgment

This research was supported/partially supported by Juwita RestiAnisa, Ananda Dwi Rahma, Dimension Data, PT, Malang Cyber Crew and ID Honeypot Project. I'm thankful to our colleagues Meilky Tandean and Iman Sayogo who provided expertise that greatly assisted the research, although they may not agree with all of the interpretations provided in this paper.

I am also grateful to Mr. d03l for assistance who inspire me to create this paper, and Michael Hendra Wijaya who moderated this paper and in that line improved the manuscript significantly.

I have to express out appreciation to the Lim, Charles for sharing their pearls of wisdom with me during the course of this research in WhatsApp Group, and apologize if there is a name that is not already listed in my paper.

References

[1] Konsep Big Data, source : http://datascience.or.id/2015/08/15/konsep-big-data/

[2] Rustock botnet, source : https://en.wikipedia.org/wiki/Rustock_botnet

[3] Video porno di videotron: Akibat kelalaian operator?, source : http://www.bbc.com/indonesia/berita_indonesia/2016/10/161005_indonesia_videotron_hack

[4] Data Theft, source : https://en.wikipedia.org/wiki/Data_theft

[5] Domain Name System, source : https://en.wikipedia.org/wiki/Domain_Name_System

[6] Bab 3 : Sistem keamanan informasi & teknik pencuria data, source :

http://panjidarmawan619.blogspot.co.id/2013/11/bab-3-sistem-keamanan-informasi-teknik.html

[7] Mengenal Serangan Man-in-The-Middle (MITM), source : http://www.ilmuhacking.com/basic-concept/mengenal-serangan-man-in-the-middle-mitm/

[8] Chapter 1. Introduction, source : https://www.wireshark.org/docs/wsug_html_chunked/ChapterIntroduction.html

[9] HTTP - HyperText Transfer Protocol, source : http://www.webopedia.com/TERM/H/HTTP.html

[10] Postman - Modern software is built on APIs, source : https://www.getpostman.com