Measuring Privacy Threats in China-Wide Mobile Networks

Mingming Zhang*, Baojun Liu*, Chaoyi Lu*, Jia Zhang*✉, Shuang Hao† and Haixin Duan*

* Tsinghua University, † University of Texas at Dallas

Abstract

HTTP transparent proxies are widely deployed in mobile networks and can lead to potential security and privacy issues. HTTP traffic is increasingly subject to in-path manipulation, especially in cellular networks. Although the traffic manipulation behavior has been studied for long, little has been understood about the manipulation and privacy concerns that arise on networks in China, due to the limitations in measurement vantage points. In this work, we aim to fill this research gap by collecting large-scale HTTP sessions originating from China-wide mobile networks, and investigating potential privacy threats caused by HTTP transparent proxy devices. Our findings are multi-faceted. First, contents of web pages can be modified by proxy devices, which are replaced by or injected with advertisements. Second, HTTP headers with user-related and device-related data are injected into HTTP requests, which raises privacy concerns. In particular, we also find HTTP headers that embed exploit codes. Our study sheds light on the HTTP traffic manipulation behavior in China-wide mobile networks, and discusses related privacy threats.

1 Introduction

Transparent proxy servers, including NAT devices, cache servers, enterprise proxies, firewalls and other devices, are increasingly deployed by mobile network operators, in order to enhance network performance and security [10]. A prior long-study of devices that manipulate HTTP traffic uncovered how the devices lead to numerous privacy and security vulnerabilities [16]. Other studies have shown that, it is possible for operators of mobile networks to track users by injecting identifiers into HTTP headers [14], or monetize by injecting advertisements into web pages [1, 8]. Moreover, some transparent proxies are vulnerable to known attacks, which threatens the privacy of all users in the same network [15].

However, due to the geographical limitations of vantage points, HTTP traffic manipulation in China-wide mobile networks has not been comprehensively studied. To start addressing this gap, in this work we investigate the privacy threats posed by HTTP transparent proxies in cellular networks, mainly from the vantage point of China.

Our Study. In this paper, we perform a comprehensive measurement study on HTTP transparent proxies in China-wide mobile networks with a focus on traffic manipulation behaviors. We collaborate with a leading security company which provides network debugging tools for millions of active mobile users. Leveraging their tool, we collect HTTP traffic between our controlled website and mobile clients located in China. Then, by checking whether HTTP sessions are manipulated in their path (e.g., modified HTML web pages or injected HTTP headers), we identify the presence of transparent proxies. Furthermore, through clustering the headers and web pages, we classify the manipulation behaviors and explore their motivations.

Findings. In this study, we develop the following key findings. In total, we collect 33,439 HTTP sessions from 30K cellular IP addresses over China in two weeks, and find 3.86% (1,291 sessions) of our collected traffic is manipulated by proxy devices. Geographically, manipulated traffic is found in 30 provinces (of 34 in total) and 16 network providers, which include large operators such as China Telecom, China Unicom and China Mobile. Regarding manipulation behaviors, we find both modified HTML web pages and injected HTTP headers. First, 22 web pages in our dataset are modified for purposes including traffic monetizing. Second, 43 types of HTTP headers embedded with privacy information are injected into 1,271 sessions, which can be used to identify or track mobile users (e.g., `x-up-calling-line-id` and `x-source-id`). In particular, we find two malicious traffic manipulation cases, where payloads exploiting software vulnerability are injected into HTTP headers.

Roadmap. The remainder of this paper is organized as follows. We first summarize previous related work in Section 2. Section 3 illustrates our methodology, including how we collect traffic from China-wide mobile networks and how we identify manipulated traffic. Section 4 presents our findings and analysis regarding privacy threats. And finally, Section 5 concludes the paper.

2 Related Work

Understanding security and privacy issues of HTTP middleboxes in networks has become a hot topic in recent years. Most early works focus on detecting and comprehensively understanding the roles of middleboxes [4, 10, 16, 20]. Network measurement tools, such as Netalyzr [7], have been developed to provide different kinds of network function tests towards middleboxes for end users. Furthermore, characteristics of HTTP proxies and DNS proxies in cellular networks have been studied based on their collected sessions [7, 17].

As a result, a number of emerging security and privacy threats related to HTTP proxies have been discovered. For instance, network operators have been found to inject user- and device-related information into HTTP headers for advertising purposes [13–15]. Others have been found to embed advertisements that load malicious code or malware into web content [1, 8]. Different kinds of HTTP header enrichment in cellular networks are studied by Netalyzr, which is most related to our work. However, Chinese networks are not included in their dataset. Recently, Hola, a peer-to-peer network has been leveraged to collect measurement dataset, detect HTTP middleboxes and understand their behaviors in the wild [1, 5, 13]. However, the studies do not focus on mobile networks.

Compared to previous research, our work gives a large-scale measurement on HTTP traffic manipulation in China-wide mobile networks, serving as a complement to prior works.

3 Methodology

To comprehensively understand HTTP traffic manipulation and associated privacy threats, we first need to obtain a large volume of vantage points in mobile networks. Subsequently, native HTTP traffic datasets are collected from the clients to identify manipulation behaviors. In this section, we elaborate our methodology of collecting the dataset and identifying manipulation.

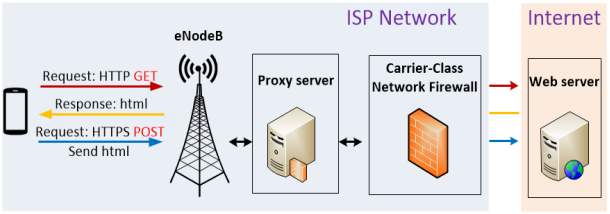

Figure 1: *Workflow of data collection*

3.1 Dataset Collection

Our collected dataset includes HTTP traffic between a controlled website and clients distributed in China-wide mobile networks. We cooperate with our industrial partner and embed our measurement code into a mobile network debugging tool, which enables the data collection process.

We control the content hosted on our web server and use it to identify traffic modification by comparing our original content to that received by the mobile clients. As shown in Figure 1, when the tool is manually run by a mobile user, it sends an HTTP `GET` request to our server to fetch its web page. After receiving the response, the tool compares the web page to our original one by calculating their MD5 hash values. If there is a mismatch (i.e., page has been modified), the modified page is sent back to our server through a `POST` request. In order to protect privacy of mobile users, web contents are encrypted before being sent back.

Collected dataset. After filtering invalid traffic such as malformed packets, we collect a total of 33,439 HTTP sessions from 30,810 mobile IP addresses in China. The clients span 31 (of 34) Chinese provinces, with Guangdong (11.0%), Beijing (9.9%) and Jiangsu (7.2%) topping the list. Zooming into AS level, our clients are distributed in 79 ASes, with 11,206 (35.9%) addresses belonging to 15 ASes of CMNET (China Mobile Infrastructure Network, e.g., AS4134 and AS9808).

Limitation. The major limitation of our dataset collection concentrates on whether a client is connected to a cellular network or a wired network like Wi-Fi. As we do not have control over the clients or read their network status (nor should we), it is difficult to perform a filtering merely based on their HTTP traffic of a website. We plan to develop possible partition features in our future work.

Ethics. Our dataset collection may raise ethical concerns, and here we discuss them explicitly. Throughout this study, we take utmost care to protect users from side-effects which may be caused by our experiment.

To collect our dataset, we embed our measurement code into a mobile network debugging tool, which is owned by a leading security company. This tool, which is similar to Netalyzr, comes with a consent stating its procedure and data collection in the user agreement upon installation, and needs to be installed and run manu-

Figure 2: *Examples of modified web pages. The page on the left provides authentication to accessing free Wi-Fi services. The page on the right contains advertisements of China Unicom for monetization purposes.*

ally by mobile users. Users reserve the right to choose whether to install or update the software, and whether to run this module with our measurement code. Therefore, our script only runs when granted with permission from the mobile user. In addition, on each client we only request our controlled web page, and only modified web pages (detected by an MD5 check) are collected, in order to mitigate privacy issues[1].

To avoid privacy concerns regarding our collected dataset, traffic from clients is encrypted before being sent back to our server. In addition, our collected dataset is securely stored in an encrypted disk of our industrial partner, and is exclusively accessible for researchers of this paper. Through said approaches, we believe we have minimized the concerns regarding users' privacy and security in the experiments.

Mobile users typically are not aware of HTTP transparent proxies or the privacy threats being introduced. By participating in our study, users could notice the problem, as our tool is able to check whether the web page is modified by calculating hashes. In the long term, our study helps understanding and protecting the privacy of mobile users.

3.2 Identifying Manipulation

Our main idea to identify transparent proxies, is checking whether HTTP traffic between clients and our website is manipulated. Here, traffic manipulation includes modification of HTML web contents (e.g., new HTML DOM tags and page replacement, examples given in Figure 2) and HTTP headers (e.g., injected HTTP headers). Since we control the website and its original content, traffic manipulation can be detected by simply comparing contents received by mobile clients.

In practice, Figure 3 presents the infrastructure of transparent proxy identification. HTTP headers and HTML web pages are separately extracted from traffic that reaches our server. For HTTP headers, we calculate

[1]To keep anonymity of our partner company as requested, here we do not provide graphic details of the software.

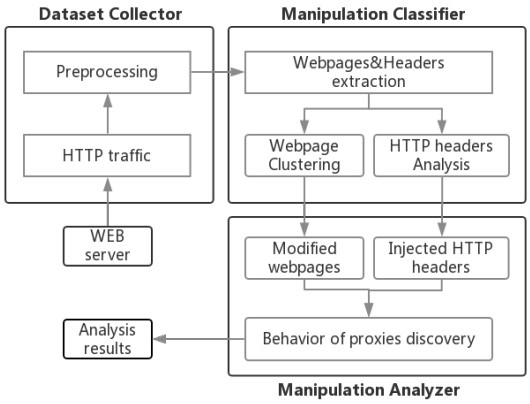

Figure 3: *Identifying HTTP transparent proxies*

the Jaccard distance

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

between the original header set and captured header set to measure their difference.

For HTML web pages, unsupervised hierarchical clustering is used to group similar web pages, in order to classify how they are modified. To this end, each web page is first formatted and converted to a vector of four features, including (1) *HTML length*: length of the entire HTML document; (2) *HTML tag difference*: `Jaccard` similarity coefficient between HTML tags of web pages; (3) *page title difference*: editing distance of the `title` field; and (4) *DOM tree difference*: minimum steps of transforming one tree to another. Subsequently, we calculate the Euclidean distances between the vectors, as a variable of web page similarity. And finally, by manually inspecting sample pages from each cluster, types of manipulation can be quickly confirmed.

4 Analysis

In this section, we present our measurement findings of transparent proxies in China-wide mobile networks. We begin by describing its scale, followed by analysis on how HTTP traffic can be manipulated by the in-path devices.

4.1 Scale of Traffic Manipulation

Among our 33K collected sessions in China-wide mobile network, we find 1,291 HTTP sessions (3.86%) from 451 (1.46%) clients are manipulated by proxy devices. In 1,271 sessions, HTTP headers are modified. Particularly, we find that privacy data can be embedded into injected headers, such as phone numbers and IMEI. As for HTML contents, 22 web pages are found to be re-

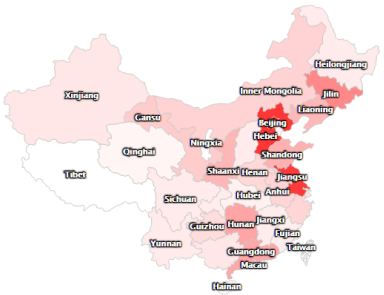

Figure 4: *Geo-distribution of mobile clients that receive manipulated traffic in China.*

placed by other contents, or injected with HTML tags by transparent proxies.

Geo- and AS-distribution. Figure 4 presents the geo-distribution of mobile clients whose HTTP traffic is manipulated. The 451 client IP addresses span 30 (of 34) provinces of China, and mainly locate in eastern provinces. For HTTP sessions, as shown in Table 1, Beijing, Hebei and Jiangsu account for most manipulated traffic. As for operators, illustrated in Table 2, most proxied sessions originate from ASes of China Telecom, China Unicom and China Mobile, which are three major network operators in China with cellular network infrastructure. The large operators own a large volume of IPv4 addresses [12] and have the most subscribers [2], thus they account for major manipulated sessions.

Table 1: *Top 8 provinces with manipulated HTTP sessions*

Province	BJ	HB	JS	JL	HN	SD	GD	SX
# Session	229	135	135	75	69	67	46	44

Table 2: *Top 5 ASes with manipulated HTTP sessions*

AS	**# Session**	**ISP**
4134	257 (19.9%)	China Telecom
4837	202 (15.6%)	China Unicom
9808	128 (9.9%)	China Mobile (GD)
4808	114 (8.8%)	China Unicom (BJ)
56046	111 (8.6%)	China Mobile (GD)

4.2 Modification of HTML Contents

Among our collected traffic, 22 HTML web pages are modified, which are either replaced by another web page or injected with new HTML tags. Below, we elaborate on the modified web contents received by clients. As the domain name of our web site is self-registered, the scale of HTTP traffic manipulation could only provide a lower bound.

Webpage replacement. 10 of 22 web pages are completely replaced by, or redirected to other contents by proxy devices. Examples of modified pages are provided in Figure 2. After manual inspection, we find that 4 pages are authentication websites for network access, such as the example on the left in Figure 2 which provides free Wi-Fi service, and public network access provided by companies such as airports. Advertisements (e.g., for wireless modems) are embedded in those pages. On the other hand, the remaining 6 pages are advertisements that promote services from the network providers and Internet companies.

HTML tag injection. Besides replaced ones, 12 other HTML documents are injected with new tags, including `<javascript>`, `<iframe>` and `<a href=URL>`. Below, we analyze their behaviors.

- *Web page rewriting.* Using JavaScript code (e.g., `document.write()`), rewriting the web page or dynamically loading extra content is possible. As examples, 2 web pages are injected with JavaScript code, which loads extra floating advertisements on the side of the pages.
- *Web page redirecting.* We find two approaches where users can be directed to other websites instead of the original one. Firstly, tags with `href` can be injected, which open a specified URL upon click. If the additional link is clicked by mobile user, other websites could be shown. For example, an extra link is found in one web page, which points to `wifi.wiair.com` and requests an HTML resource, showing classified advertisements. Secondly, leveraging JavaScript code (e.g., `window.location.href`), the redirection process can be automatic. For example, a redirection to `m.rong360.com` is conducted by injected JavaScript, which provides online loan services.
- *Resource requesting.* Since the `src` attribute in `<script>` tag points to a `JavaScript` file, it's feasible to request a new URL originated from another website. We find 2 cases where the injected `src` attributes request a URL under `www.wayos.net`, hosted by WayOS which is a networking service provider in Guangdong province of China. Besides the domain name, the URL is embedded with device information such as device name, IP address and MAC address.

Who is behind the modification? We manually inspect each modified web page, in order to investigate their modification purposes. As a result, we classify them into the following 3 categories.

- *Advertising*: Six web pages are modified to promote products and services of mobile operators (e.g., roaming service and phone card), originating from `zhengzhou-airport.wiportal.cn` and `nfdnserror1.wo.com.cn`. The domain names belong to Alibaba and China Unicom, respectively. We

also find a page containing pop-up ads and banners of online shopping products that originate from `cdn.wiair.com` and provide public Wi-Fi service, followed by a redirection to `m.baidu.com` which provides advertisements upon click. Additionally, one page injected with JavaScript is found to visit `m.rong360.com` and open a new web page with loan services.

- *Authentication*: We find 6 modified pages that are login pages or gateways for network access, and include advertisements of Internet Content Providers (ICP). For example, `zhengzhou-airport.wiportal.cn` promotes wireless modems while providing authentication.
- *Others*: There are two fake authentication pages guiding users to input their phone numbers and passwords. Others only show 404 Error Page.

In this study, we examined how HTTP contents are modified in-flight, finding that the main purpose of manipulation is promoting products or services for monetizing traffic. Our findings complement results of previous works that study this problem from out-of-band injection [8], and in-line with recent research [1], which find 0.95% exit nodes receive modified HTML pages on wired networks. In addition, compared to similar experiments in other countries, e.g., 1.9% of all US clients are affected by content modification [19], our results show a lower ratio in China-wide mobile networks (491 of 30,810 clients, 1.6%).

4.3 Modification of HTTP Headers

Compared to web pages, we find that HTTP headers are modified in more sessions. In total, 1,271 HTTP sessions in our dataset are injected with 43 types of headers. Among the injected headers, 8 types are ones that are often used by browsers, such as `Accept`, `Content-Length` and `Content-Type`.

Table 3 presents examples of injected headers, which are found to *embed privacy data of mobile users or devices*, such as location, IP address and device serial numbers (e.g., IMEI). For instance, several extension headers (e.g., `x-IMEI, x-Nx-apn and x-up-bear-type`) related to users are found in networks of large providers, including China Mobile, China Telecom and China Unicom.

However, we also find some injected headers that contain other device vendors instead of network operators (e.g., `x-huawei-apn`), thus we cannot accurately distinguish whether the injection is conducted by network operators. Therefore, we only claim that the modification is found in the networks, but not necessarily conducted by their operators, due to our methodological limitations. Below, we classify injected headers into 3 categories according to the embedded data, and separately discuss their prevalence and motivation.

4.3.1 Headers to Identify Users

Among all injected headers, 11 kinds embed user-identification data. The first part of Table 3 shows examples of injected HTTP headers (e.g., `x-IMEI`) which reveal the identity of mobile users. As HTTP requests are transfered in plain text, privacy data in the injected headers could be snooped by eavesdroppers, putting mobile users at risk.

IMEI. The International Mobile Equipment Identity (IMEI) is used to uniquely identify each mobile device [3]. IMEI could be used to identify mobile users and record their behaviors. We find 12 HTTP sessions are injected with `x-IMEI` header, which contain the IMEI of mobile devices.

IMSI. International Mobile Subscriber Identification (IMSI) is stored in the SIM card, and is unique across the global network [18]. As mobile operators or third parties are able to locate mobile devices by checking the value of IMSI, embedding IMSI in HTTP headers brings privacy threats to mobile users. In total, we find 5 HTTP sessions are injected with headers containing IMSI.

Phone Number. Headers like `x-up-calling-line-id` contain phone numbers of mobile users, which should be treated as private. Users could receive spam messages or unwanted calls if their phone numbers are leaked to malicious parties. In our collected dataset, 50 sessions leak phone numbers of their subscribers, exposing users to privacy threats.

Type of mobile service. `x-huawei-NetworkType` and `x-up-bear-type` are two headers which specify the type of mobile communication service, such as CDMA and GSM. Service types differ among network providers, e.g., China Unicom offers GSM and WCDMA services, while China Mobile offers GSM and TD-SCDMA. Therefore, type of mobile service can be used to distinguish the provider where a client resides, and to obtain network configurations. As shown in Table 3, mobile service information is injected into 128 HTTP sessions.

As discussed above, diverse user-related data embedded in HTTP headers can be used to identify mobile users, introducing privacy threats. Network operators, websites, ad providers as well as adversaries can uniquely identify the individuals by these headers. Once leaked, identification data can be used in malicious behaviors including spam.

Table 3: *Extension-headers used to identify and track mobile users*

Header Name	Type	Count	Operator (Province)
`x-IMEI`*	IMEI	12	China Mobile (GD)
`x-huawei-IMSI`*	IMSI	6	China Telecom, China Unicom
`x-up-calling-line-id`	Phone number	50	China Telecom (SH, SN, QH, SC, XJ, GS, BJ, SD, LN, YN, NM, ZJ, AH), China Mobile (GD), China Unicom (BJ, JL, LN)
`X-Nokia-CONNECTION-MODE`	Connecting mode	11	China Mobile (GD)
`x-up-bear-type`	Type of mobile service	122	China Mobile (GD), China Telecom (BJ, SH, SX, QH, SC, XJ, GS, YN), China Unicom (BJ, NM)
`x-huawei-NetworkType`*	Type of mobile service	6	China Unicom, China Telecom
`X-Forwarded-For`	Client IP	139	Farahoosh, China Mobile (GD, SD), China Telecom (SH, SX, SC, QH, XJ, GS), PT Telkom, China Unicom (JL, LN, XJ)
`X-Nx_remoteip`*	Client IP	3	China Telecom (QH, SC)
`X-Clientport`*	Port number	1	China Telecom (QH)
`X-Nokia-gateway-id`	Gateway ID	11	China Mobile (GD)
`Via`	Proxy	59	Farahoosh, China Telecom (SN, GS, YN, NM), Cybertel Telecom, China Unicom (NM, JL, XJ, LN), China Mobile (GD), INET
`x-huawei-NASIP`*	Gateway Info.	5	China Unicom
`x-source-id`	Gateway Info.	62	China Unicom (JL, LN), China Mobile (GD), China Telecom (SH, SN, QH, SC, XJ, YN, NM, JS)
`Cdn-Src-Ip`*	Client IP.	24	CNISP-Union Technology (BJ) , China Unicom (LN)
`x-huawei-apn`*	APN	5	China Unicom
`x-Nx-apn`*	APN	15	China Mobile (GD), China Telecom (JS)

* These injected HTTP headers have not been discovered in previous studies.

4.3.2 Headers to Track Users

Besides headers with identification, 9 kinds of headers can be used to track mobile users, as shown in the second part of Table 3. Operators could track mobile users by adding HTTP headers as identifiers and provide user-tracking data to advertisement platforms.

Client IP address. The `X-Forwarded-For` header contains client IP addresses [6] and is widely used by proxy servers, which is a common approach for a web server to identify the source IP address of a client. In addition, client IP addresses can be embedded in other headers, such as `X-Nx_remoteip`. We find `X-Forwarded-For` in 139 HTTP requests and `X-Nx_remoteip` in 3 requests.

Gateway Configuration. Configuration of the GPRS Support Node contains location and version data. For instance, `x-huawei-NASIP` carries the IP of NAS (Network Access Server), which is a server enabling an ISP to provide Internet access to users. Another example, `X-Nokia-gateway-id` exposes the version of Nokia's WAP gateway. As a result, the headers can be used to track subscribers of these device manufacturers.

4.3.3 Special Types

In particular, we find HTTP headers which are speculated to be injected by compromised HTTP proxy devices. The `Content-Type` fields of two HTTP sessions contain ONGL codes, which are probes for a vulnerability of Apache Struts2 (CVE-2017-5638) [9]. As an open-source web application framework for developing Java EE web applications, Struts2 parses and checks the `Content-type` field of every incoming HTTP request.

Figure 5: *Web pages of Referrer links. Both links point to Baidu search results of a keyword. The first keyword (SK8) matches a flight number and a brand for shoes. The second (ST1) matches Zenvos supercar and a game named World of Tanks.*

Upon receiving an invalid value, while throwing an exception [11], the exploit code starts to run, trying to execute commands on the host system. After manual inspection, we find that the injected code in the 2 sessions tries to execute `whoami` and `nMaskCustomMuttMoloz` commands. If remote execution is successful, attackers could obtain higher system privileges [9]. Interestingly, the two sessions originate from different ASes and network operators.

In addition, proxies can *rewrite* the values of standard headers, such as `Referer`. We find two sessions whose `Referer` are embedded with URLs pointing to search result pages (e.g., `https://www.baidu.com/s?wd=ST1`, where ST1 is the search keyword), as shown in Figure 5. However, the exact motivation behind embedding links to search result pages remains uncertain.

Table 4: *Top 15 operators related to HTTP traffic manipulation. 90% manipulated traffic are found in networks of the top 3 ISPs.*

ISP	# Session	ISP	# Session	ISP	# Session
China Mobile	524	Guoxin bilin Telecom	4	UCloud	1
China Unicom	325	Founder Broadband	3	flash newsletter cas	1
China Telecom	317	Anchang Network Security	2	New-Billion Telecom	1
CNISP-Union	15	GIANT	2	Yiantianxia Network	1
Zhejiang Taobao	8	China TieTong	1	Datong Coal	1

4.4 Motivations

In this section, based on our measurement results, we aim to investigate motivations behind HTTP traffic manipulation behaviors.

Advertising. As discussed in Section 4.2, various advertisements such as shopping ads, are injected into HTML web pages received by clients. Specifically, 6 of 22 manipulated web pages are promoting services or products which are related to mobile network operators, including China Unicom and Billion Connect. In addition, HTTP headers are used to identify users and embed information of networking services. Although the exact location of manipulation behaviors remains uncertain due to our methodological limitations, we regard that the manipulations are conducted by mobile operators or ISPs for revenue, based on the modified web contents which are displayed to mobile users.

Malicious behaviors. We also find malicious behaviors in our measurement results. Firstly, we find that exploit code probing vulnerability is injected into HTTP headers. In addition, two fake Internet authentication pages requiring users to input phone numbers and password are found. The malicious behaviors could be used to steal user privacy data or credentials.

User tracking. Proxy devices can inject headers which embed unique identifiers into HTTP requests. The headers can be used by commercial companies to identify and track the actions of mobile users. Subsequently, advertisers may send ads to their potential customers, according to users' behavior.

5 Conclusion

In order to enhance network performance and security, service providers increasingly deploy transparent proxy servers. In this paper, we perform a large-scale measurement study on HTTP traffic manipulation in China-wide mobile networks. From our collected dataset, we present a detailed investigation into diverse manipulation behaviors. We find 3.86% of our collected HTTP traffic is modified by transparent devices. What makes the behavior more problematic is that user-related and device-related headers are found to be embedded into HTTP requests, which raises privacy concerns. Also of great concern are the few cases of exploit code distributed through the proxies. Our study sheds light on the HTTP traffic manipulation and privacy threats in China-wide mobile networks.

As our future work, we plan to study the exact location of traffic manipulation, by leveraging TTL-limited requests. Further, we intend to distinguish in-path injections from on-path ones. We are also interested in studying TCP/IP level anomalies introduced by content or header modifications.

Acknowledgments

We sincerely thank our shepherd Seda Guerses and all anonymous reviewers for their insigtvauable suggestions and comments to improve the paper. We also thank Yishuai Liu, Deliang Chang, Yuchuan Hu, Mingxuan Liu and Lu Liu for their feedback and help.

This work was funded by the National Natural Science Foundation of China (grant #61472215 and #U1636204). Any views, opinions, findings, recommendations, or conclusions contained or expressed herein are those of the authors, and do not necessarily reflect the position, official policies or endorsements, either expressed or implied, of the Government of China.

References

[1] CHUNG, T., CHOFFNES, D., AND MISLOVE, A. Tunneling for transparency: A large-scale analysis of end-to-end violations in the internet. In *Proceedings of the 2016 ACM on Internet Measurement Conference* (2016), ACM, pp. 199–213.

[2] CIW. Top 3 players of chinas telecom market. https://www.chinainternetwatch.com/22766/china-telecoms-q3-2017/. [Online; accessed 15-July-2018].

[3] GSM. Imei allocation and approval guidelines. https://www.gsma.com/newsroom/wp-content/uploads/2012/06/ts0660tacallocationprocessapproved.pdf. [Online; accessed 6-July-2018].

[4] HUANG, S. *Detecting Middlebox Interference on Applications*. PhD thesis, Queen Mary University of London, 2017.

[5] HUANG, S., CUADRADO, F., AND UHLIG, S. Middleboxes in the internet: a http perspective. In *Network Traffic Measurement and Analysis Conference (TMA)* (2017), IEEE, pp. 1–9.

[6] JONES, P., PEARCE, C., GIRALT, P., AND SALGUEIRO, G. End-to-end session identification in ip-based multimedia communication networks. https://www.rfc-editor.org/info/rfc7989, October, 2016.

[7] KREIBICH, C., WEAVER, N., NECHAEV, B., AND PAXSON, V. Netalyzr: illuminating the edge network. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement* (2010), ACM, pp. 246–259.

[8] NAKIBLY, G., SCHCOLNIK, J., AND RUBIN, Y. Website-targeted false content injection by network operators. In *USENIX Security Symposium* (2016), pp. 227–244.

[9] RAPID7. Nvd cve-2017-5638 detail. https://www.rapid7.com/db/vulnerabilities/apache-struts-cve-2017-5638. [Online; accessed 8-July-2018].

[10] SHERRY, J., HASAN, S., SCOTT, C., KRISHNAMURTHY, A., RATNASAMY, S., AND SEKAR, V. Making middleboxes someone else's problem: network processing as a cloud service. *ACM SIGCOMM Computer Communication Review 42*, 4 (2012), 13–24.

[11] STRUTS. Apache struts exception handling. https://struts.apache.org/getting-started/exception-handling.html. [Online; accessed 3-July-2018].

[12] STRUTS. Statistical report on internet development in china. https://cnnic.com.cn/IDR/ReportDownloads/201706/P020170608523740585924.pdf. [Online; accessed 6-July-2018].

[13] TYSON, G., HUANG, S., CUADRADO, F., CASTRO, I., PERTA, V. C., SATHIASEELAN, A., AND UHLIG, S. Exploring http header manipulation in-the-wild. In *Proceedings of the 26th International Conference on World Wide Web* (2017), International World Wide Web Conferences Steering Committee, pp. 451–458.

[14] VALLINA-RODRIGUEZ, N., SUNDARESAN, S., KREIBICH, C., AND PAXSON, V. Header enrichment or isp enrichment?: Emerging privacy threats in mobile networks. In *Proceedings of the 2015 ACM SIGCOMM Workshop on Hot Topics in Middleboxes and Network Function Virtualization* (2015), ACM, pp. 25–30.

[15] VALLINA-RODRIGUEZ, N., SUNDARESAN, S., KREIBICH, C., WEAVER, N., AND PAXSON, V. Beyond the radio: Illuminating the higher layers of mobile networks. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (2015), ACM, pp. 375–387.

[16] WEAVER, N., KREIBICH, C., DAM, M., AND PAXSON, V. Here be web proxies. In *International Conference on Passive and Active Network Measurement* (2014), Springer, pp. 183–192.

[17] WEAVER, N., KREIBICH, C., NECHAEV, B., AND PAXSON, V. Implications of netalyzrs dns measurements. In *Proceedings of the First Workshop on Securing and Trusting Internet Names (SATIN), Teddington, United Kingdom* (2011).

[18] WIKIPEDIA CONTRIBUTORS. International mobile subscriber identity. https://en.wikipedia.org/wiki/International_mobile_subscriber_identity. [Online; accessed 15-July-2018].

[19] ZHANG, C., HUANG, C., ROSS, K. W., MALTZ, D. A., AND LI, J. Inflight modifications of content: who are the culprits? In *LEET* (2011).

[20] ZHANG, H., AND CHOFFNES, D. Client-side web proxy detection from unprivileged mobile devices. *arXiv preprint arXiv:1511.04493* (2015).