Detection for Trusted Content Delivery Networks Traffic by Pattern-based Content Leakage

Madhavi R Suryawanshi
Computer Engineering
G.H.Raisoni College of
Engineering and Management
Pune

Sarita A Patil
Computer Engineering
G.H.Raisoni College of
Engineering and Management
Pune

ABSTRACT
Due to growing reputation of multimedia systems surging applications and solutions nowadays, the issue of trusted online video supply in order to avoid undesired content leakage possesses, certainly, become essential. Even though keeping user comfort, standard systems get tackled this challenge by simply proposing methods in line with the observation of streamed traffic through the circle. Most of these standard systems keep a high prognosis precision though dealing with a few of the traffic variance inside circle (e. g.,circle hold up and bundle loss), nonetheless, the prognosis overall performance drastically degrades as a result of the particular significant variance of online video programs. Within this papers, all concentrate on defeating this challenge by simply proposing any fresh content-leakage prognosis program which is effective towards variance in the online video size. Through researching videos of diverse programs, all ascertain any regards between the duration of videos to get in comparison and the particular likeness between your in comparison videos. Consequently, enhance the prognosis overall performance in the suggested program also in the natural environment the subject of variance in length of online video. Via a test bed try, the potency of your suggested program is considered with regard to variance of online video size, hold up variance, and bundle damage.

Keywords
Streaming content, leakage detection, traffic pattern, degree of similarity.

1. INTRODUCTION
Nowadays, with all the swift advancement connected with broadband technology plus the improvement connected with high-speed wired/wireless systems, the recognition connected with real-time online video media streaming apps in addition to companies over the web has enhanced by jumps in addition to bounds. YouTube in addition to Microsoft company circle online video media usually are distinctive examples of like apps. They function a big people connected with customers from all over the world having different subject matter, between regular reports for you to activity for such as new music, video lessons, athletics, or anything else, by making use of streaming transmitting technology. Moreover, real-time online video media streaming sales and marketing communications like internet seminar in intracompany systems as well as by way of Net having exclusive systems (VPNs) will be broadly implemented in a large number of firms to be a powerful method of proficiently endorsing business activities without having extra fees.

A significant matter in online video media streaming companies will be the safeguard in the bit supply from unauthorized work with, duplication in addition to submitting. Probably the most well-liked methods to avoid unwelcome subject matter submitting for you to unauthorized customers and/or to protect authors' copyrights will be the digital legal rights managing (DRM) technological innovation. The majority of DRM approaches employ cryptographic as well as digital watermark approaches. On the other hand, these kinds of solutions don't have a substantial impact on redistribution connected with subject matter, decrypted as well as refurbished on the user-side by authorized still harmful customers. Furthermore, redistribution will be officially will no longer difficult by applying peer-to-peer (P2P) streaming software. Hence, streaming traffic may be released for you to P2P systems.

However, box selection by firewall-equipped egress nodes can be an uncomplicated strategy to steer clear of leakage connected with streaming subject matter for you to external systems. With this option, the box header data (e. gary., desired destination in addition to resource Net protocol details, protocol variety, in addition to vent number of outgoing traffic) of each and every streamed box will be checked out. Just in case the checked out packets usually do not examine the predefined selection insurance plan, they're clogged in addition to dropped. On the other hand, it truly is difficult for you to totally avoid streaming content material leakage by using box selection by itself considering that the box header data connected with harmful customers will be unspecified beforehand and may be quickly spoofed.

With this report, all of us target the unlawful redistribution connected with streaming content material by a certified individual for you to external systems. The present proposals in monitor data received from various nodes in the middle of the streaming route. The particular retrieved data utilized to generate traffic habits which usually look because special waveform each content material just like a fingerprint. The particular age group connected with traffic style doesn't demand any kind of information on the box header, and so saves the user's solitude. Loss detection will be then conducted by evaluating the made traffic habits. On the other hand, the lifetime connected with video lessons connected with various size inside the circle natural environment reasons some considerable degradation inside the leakage detection performance. Hence, building a progressive leakage detection method sturdy on the variance connected with online video media measures will be, without a doubt necessary. With this report, by evaluating different length video lessons, all of us establish a new romantic relationship concerning the size of video lessons to become compared in addition to their own likeness. Dependent about this romantic relationship, all of us establish conclusion tolerance which allows exact leakage

detection actually within an natural environment having various size video lessons.

2. RELATED WORK

In this section discuss existing work done by the researchers on Detection for trusted content delivery networks

Privacy-Preserving Detection of Sensitive Data Exposure Statistics from security firms, research institutions and government organizations show that the number of data-leak instances have grown rapidly in recent years.[1] Among various data-leak cases, human mistakes are one of the main causes of data loss. There exist solutions detecting inadvertent sensitive data leaks caused by human mistakes and to provide alerts for organizations. A common approach is to screen content in storage and transmission for exposed sensitive information. Such an approach usually requires the detection operation to be conducted in secrecy. However, this secrecy requirement is challenging to satisfy in practice, as detection servers may be compromised or outsourced. This paper, present a privacy preserving data-leak detection (DLD) solution to solve the issue where a special set of sensitive data digests is used in detection. The advantage of our method is that it enables the data owner to safely delegate the detection operation to a semihonest provider without revealing the sensitive data to the provider. System describes how Internet service providers can offer their customers DLD as an add-on service with strong privacy guarantees. The evaluation results show that our method can support accurate detection with very small number of false alarms under various data-leak scenarios. System proposed fuzzy fingerprint, a privacy-preserving data-leak detection model and present its realization. Using special digests, the exposure of the sensitive data is kept to a minimum during the detection. Conducted extensive experiments to validate the accuracy, privacy, and efficiency of our solutions. For future work, plan to focus on designing a host-assisted mechanism for the complete data-leak detection for large-scale organizations.

Quantifying information leaks in outbound web traffic present an method for quantifying data leak capability in the network traffic.[2] Instead of trying to sense the occurrence of sensitive information an impossible task in the universal case a their goal is to calculate and limit its highest volume. They also propose the measurement algorithms for the Hypertext Transfer Protocol (HTTP), the main protocol for web browsing . The results were best for the blog scenario because the blog website. The main advantage of this paper is that this paper insight that most network traffic is repeated or determined by external. In this system network traffic is so voluminous that manual inspection would be unreasonably expensive.

This paper introduced a new approach for quantifying information leaks in web traffic. Instead of inspecting a message's data, the goal was to quantify its information content. The algorithms in this paper achieve precise results by discounting fields that are repeated or constrained by the protocol. This work focuses on web traffic, but similar principles can apply to other protocols. Our analysis engine processes static fields in HTTP, HTML, and Java script to create a distribution of expected request content. It also executes dynamic scripts in an emulated browser environment to obtain complex request values.

Panorama: Capturing system-wide information flow for malware detection and analysisfor identifying and breakdown malware by catching this fundamental trait.[3] In this system extensive experiments, Panorama effectively identified all the malware tests and had not very few false positives. Besides, by utilizing Google Desktop as a case study, they demonstrate that their framework can precisely catch its data get to and preparing conduct, and they can confirm that it sends back delicate data to remote servers in specific settings. This system can accurately capture its information access and processing behavior. Malicious programs spy on users behavior and compromise their privacy. Even software from reputable vendors, such as Google Desktop and Sony DRM. Google Desktop as a case study, they show that our system can accurately capture its information access and processing behavior, and they can confirm that it does send back sensitive information to remote Servers in certain settings.

Malware has brought along serious security and privacy threats. However, existing techniques for malware detection and analysis are ineffective. This paper proposed whole-system fine-grained taint analysis to discern fine-grained information access and processing behavior of a piece of unknown code. This behavior captures the intrinsic characteristics of a wide-spectrum of malware, including key loggers, password sniffers, packet sniffers, stealth back-doors, BHO-based spyware, and rootkits. Thus, the detection and analysis relying on it cannot be easily evaded. To evaluate the effectiveness of this approach, designed and developed a system, called Panorama. In the experiments, evaluated 42 malware samples and 56 benign samples. Panorama yields zero false negative and very few false positives. Then useGoogle Desktop as a case study. Demonstrated that Panorama can accurately capture its information access and processing behavior, and confirm that it does send back sensitive information to remote servers.System such as Panorama will offer indispensable assistance to malware analysts and enable them to quickly comprehend the behavior and inner-workings of malware.

Protecting confidential data on personal computers with storage capsules, a novel methodology for securing private documents on an individual PC.[4] Storage Capsules are encoded record containers that permit a compromised machine to safely view and alter sensitive documents without malware being able to steal information. The framework accomplishes this objective by taking a checkpoint of the present framework state and disabling device output before permitting access a Storage Capsule. Composes to the Storage Capsule are then sent to a trusted module. The trusted module declassifies the Storage Capsule by re-encrypting its contents, and exports it for storage in a low integrity environment. Storage Capsules are encrypted file containers that allow a compromised machine to securely view and edit sensitive files without malware being able to steal confidential data. The main limitation Suffer huge losses if private data falls into the wrong hands. One of the primary threats to confidentiality is malicious software on personal computers.

This system introduced Storage Capsules, a new mechanism for securing files on a personal computer. Storage Capsules are similar to existing encrypted file containers, but protect sensitive data from malicious software during decryption and editing. The Capsule system provides this protection by isolating the user's primary operating system in a virtual machine. The Capsule system turns off the primary OS's device output while it is accessing confidential files, and reverts its state to a snapshot taken prior to editing when it is finished. One major benefit of Storage Capsules is that they work with current applications running on commodity operating systems.

Preventing accidental data disclosure in modern operating systemsAquifer as an arrangement structure and framework for counteracting accidental data exposure in advanced working frameworks. [5] In Aquifer, application engineers define secrecy confinements that ensure the whole client interface work process characterizing the client task. Aquifer provides protection beyond simple permission checks and allows applications to retain control of data even after it is shared. Using file permission also avoids ambiguous read-write file open masks, as well as properly propagating labels when the workflow label changes between file open and file write.

Modern operating systems have changed both the way users use software and the underlying security architecture. These two changes make accidental data disclosures easier. To address this problem, presented the Aquifer security framework that assigns host export restrictions on all data accessed as part of a UI workflow. Our key insight was that when applications in modern operating systems share data, it is part of a larger workflow to perform a user task. Each application on the UI workflow is a potential data owner, and therefore can contribute to the security restrictions. The restrictions are retained with data as it is written to storage and propagated to future UI workflows that read it. In doing so, enable applications to sensibly retain control of their data after it has been shared as part of the user's tasks.

3. IMPLEMENTATION DETAILS

In this section discussed about the proposed system in detail, and also discuss the system overview in detail, proposed algorithm, mathematical model of the proposed system,

3.1 System Overview

The following figure 1 shows the architectural view of the proposed system. The description of the system is as follows:

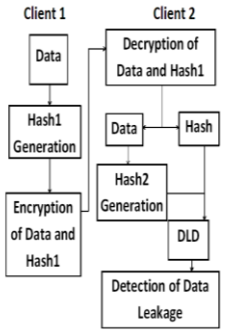

Figure 1:System Architecture

3.2 Algorithm

In this section discuss the algorithm of the proposed system, AES algorithm and SHA1 algorithm used for the hashing technique.

Algorithm 1: Proposed System Algorithm

Step 1: Generate fingerprint f, which is representation signature of data.

Step 2: $F = \{f_1, f_2, f_3, \ldots\ldots, f_n\}$ as per data.

Step 3: f' is fingerprint i.e. generated at DLD side or server side.

Step 4: $f' = \{f'_1, f'_2, f'_3, \ldots\ldots, f'_n\}$ as per data at DLD

Step 5: Compare F with F' and if both are equal then send data to destination else discard data.

Algorithm 2: AES Algorithm

The algorithms used in AES are so easy that they can be easily implemented using cheap processors and a minimum amount of memory. Very efficient Implementation was a key factor in its selection as the AES cipher. The steps involved in AES is

1. Key Expansion: - Using Rijndael's key schedule Round keys are derived from the cipher key.
2. If DistanceToTree(u) >DistanceToTree(DCM) and First-Sending(u) then
3. Initial Round :-AddRoundKey where Each byte of the state is combined with the round key using bitwise xor.
4. Rounds SubBytes : non-linear substitution step ShiftRows : transposition step MixColumns : mixing operation of each column. AddRoundKey
5. Final Round: It contain SubBytes, ShiftRows and Ad-dRoundKey

Algorithm 3: SHA1 Algorithm

Step 1: Append Padding Bits

Message is "padded" with a 1 and as many 0's as necessary to bring the message length to 64 bits fewer than an even multiple of 512.

Step 2: Append Length

64 bits are appended to the end of the padded message. These bits hold the binary format of 64 bits indicating the length of the original message.

3.3 Mathematical Model

Input

Browse Dataset

U= {u1, u2, u3,....,un}

Where, U is a set of input data and u1, u2, u3,....,un are the different dataset.

Process

Client

C = {C1, C2, C3,....,Cn}

Where, C is represented as a set of Client and C1, C2, C3,….,Cn are the number of Clients.

Bloom Filter

B= { b1, b2, b3,…..,.bn}

Where, B is represent as a set of Bloom Filter Model and b1, b2, b3,…,.bn are set of Bloom Filter Model.

Server S= { s1, s2, s3,….,sn}

Where, S is represent as a set of Server and s1, s2, s3,….,sn are number of server.

Output

Final Result FR= {fr1, fr2, fr3,…..,frn}

Where, FR is represent as a set of final result and fr1, fr2, fr3,....,frn are number of final output.

3.4 Experimental Setup

The system is built using Java framework (version jdk 8) on Windows platform. The Netbeans (version 8.1) is used as a development tool. The system doesn't require any specific hardware to run, any standard machine is capable of running the application.

4. RESULT AND DISCUSSION

4.1 DataSet

Dataset used in this system is the text file.

4.2 Results

In table 1 shows the time required for implementing the different methods of the proposed system. Time measure for different methods, the distinct methods are fingerprint encryption, encryption of data, sending the data, digital envelope, compare data fingerprint.

Table 1: Time Measurement

	Fingerprint Generation	Encryption of data	Sending the Data	Digital Envelope	Compare data fingerprint
D 1	1.08 sec	2,71 sec	0.74 sec	0.32 sec	0.087 sec
D 2	1.38 sec	3,18 sec	0.83 sec	0.37 sec	0.932 sec

The following figure shows the graph of the proposed system. Plot the graph from the above table.

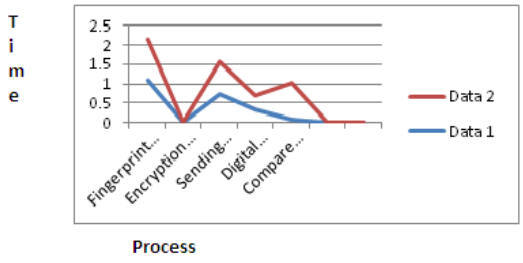

Figure 2: Time Graph

5. CONCLUSION AND FUTURE SCOPE

The content leakage detection system based on the fact that each streaming content has a unique traffic pattern is an innovative solution to prevent illegal redistribution of contents by a regular, yet malicious user. This system attempts to solve these issues by introducing a dynamic leakage detection scheme. Moreover, in this system investigate the performance of the proposed method under a real network environment with videos of different lengths.

6. ACKNOWLEDGMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. Also thankful to the reviewer for their valuable suggestions. Also thank the college authorities for providing the required infrastructure and support. Finally, would like to extend a heartfelt gratitude to friends and family members.

7. REFERENCES

[1] XiaokuiShu, Danfeng Yao, Member, and Elisa Bertino, ``Privacy-Preserving Detection of Sensitive Data Exposure", *In IEEE transactions on information forensics and security*, vol. 10, no. 5, May 2015.

[2] K. Borders and A. Parkas, "Quantifying information leaks in outbound web traffic," *in Proc. 30th IEEE Sump. Secure. Privacy*, May 2009, pp. 129-140.

[3] H. Yin, D. Song, M. Agile, C. Kruegel, and E. Kirda, "Panorama: Capturing system wide information flow for malware detection and analysis," *in Proc. 14th ACM Conf. Compute. Commun.*Secur., 2007,pp. 116-127.

[4] K. Borders, E. V. Weele, B. Lau, and A. Prakash, "Protecting confidential data on personal computers with storage capsules," in Proc. 18th USENIX Secur. Symp., 2009, pp. 367-382.

[5] A. Nadkarni and W. Neck, "Preventing accidental data disclosure in modern operating systems," in Proc. 20th ACM Conf. Compute.Commun.Secur., 2013, pp. 1029-1042.

[6] Shengbao Wang, Zhenfu Cao, Maurizio Adriano Strangio and Lihua Wang, `` Cryptanalysis and Improvement of an Elliptic Curve Diffie-Hellman Key Agreement Protocol", In IEEE Transaction on December 14, 2007.

[7] S. Jha, L. Kruger, and V. Shmatikov, "Towards practical privacy for genomic computation," in Proc. 29th IEEE Symp.Secur. Privacy, May 2008, pp. 216–230.

[8] Y. Chu, S.G. Rao, S. Seshan, and H. Zhang, "EnablingConferencing Applications on the Internet Using an OverlayMulticast Architecture," Proc. ACM SIGCOMM, pp. 55-67,Aug. 2001.

[9] Z. Yang, H. Ma, and J. Zhang, "A Dynamic Scalable Service Modelfor SIP-Based Video Conference," Proc. Ninth Int'l Conf. ComputerSupported Cooperative Work in DE, pp. 594-599, May 2005.

[10] Y. Chu, S.G. Rao, S. Seshan, and H. Zhang, "Enabling ConferencingApplications on the Internet Using an Overlay MulticastArchitecture," Proc. ACM SIGCOMM, pp. 55-67, Aug. 2001.

[11] O. Adeyinka, "Analysis of IPSec VPNs Performance in a MultimediaEnvironment," Proc. Fourth Int'l Conf. Intelligent Environments,pp. 25-30, 2008.

[12] E.I. Lin, A.M. Eskicioglu, R.L. Lagendijk, and E.J. Delp, "Advancesin Digital Video Content Protection," Proc. IEEE, vol. 93, no. 1,pp. 171-183, Jan. 2005.

[13] S. Craver, N. Memon, B.L. Yeo, and M.M. Yeung, "ResolvingRightful Ownerships with Invisible Watermarking Techniques:Limitations, Attacks, and Implications," IEEE J. Selected AreasComm., vol. 16, no. 4, pp. 573-586, May 1998.

[14] K. Xu, D. Yao, Q. Ma, and A. Crowell, "Detecting infection onsetwith behavior-based policies," in *Proc. 5th Int. Conf. Netw. Syst. Secur.*,Sep. 2011, pp. 57–64.

[15] M. O. Rabin, "Fingerprinting by random polynomials," Dept. Math.,Hebrew Univ. Jerusalem, Jerusalem, Israel, Tech. Rep. TR-15-81, 1981.

[16] A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher,"Min-wise independent permutations," *J. Comput. Syst. Sci.*, vol. 60,no. 3, pp. 630–659, 2000.

[17] A. Z. Broder, "Some applications of Rabin's fingerprinting method," in*Sequences II*. New York, NY, USA: Springer-Verlag, 1993, pp. 143–152.

[18] A. Z. Broder, "Identifying and filtering near-duplicate documents," in*Proc. 11th Annu. Symp.Combinat. Pattern Matching*, 2000, pp. 1–10.

[19] A. Broder and M. Mitzenmacher, "Network applications of bloom filters:A survey," *Internet Math.*, vol. 1, no. 4, pp. 485–509, 2004.

[20] G. Aggarwal*et al.*, "Anonymizing tables," in *Proc. 10th Int. Conf.Database Theory*, 2005, pp. 246–258.

[21] R. Chen, B. C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang,"Privacy-preserving trajectory data publishing by local suppression," *Inf.Sci.*, vol. 231, pp. 83–97, May 2013.

[22] M. O. Rabin, "Digitalized signatures and public-key functions asintractable as factorization," Massachusetts Inst. Technol., Cambridge,MA, USA, Tech. Rep. MIT/LCS/TR-212, 1979.

[23] F. Liu, X. Shu, D. Yao, and A. R. Butt, "Privacy-preserving scanningof big content for sensitive data exposure with MapReduce," in *Proc.ACM CODASPY*, 2015.

[24] W. N. Francis and H. Kucera, "Brown corpus manual," 1979.

[25] J. Kleinberg, C. H. Papadimitriou, and P. Raghavan, "On the value ofprivate information," in *Proc. 8th Conf. Theoretical Aspects RationalityKnowl.*, 2001, pp. 249–257.