ISSN:2348-2079

Volume-6 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

Cloud Resource based Data Origin Detection with Security Features

[1]Remya M. K, [2]Nikhila A,
[1]PG Scholar,[2]Assistant Professor
[1,2] Dept. of M.Tech (CSE),Cochin College of Engineering and Technology, Valiyaparambu, Edayur P.O.,Valanchery, Malappuram District, Kerala, India
E-mail id: remyajdt@gmail.com

Abstract

The network data transmission operations are handled with the Internet Protocol (IP). Each node in the network is allocated with the static or dynamic IP address values. The web applications can be disrupted from various sources. The IP traceback operations are carried out to discover the origin of the received data elements. The Internet Service Provider (ISP) maintains the logs with user data transmission details. The data origin detection is carried out with traceback log analysis. The topology structure of the Internet Service Providers can be extracted by the intruders during the IP traceback operations.

The cloud resources are utilized in the IP traceback operations. The authentication and security are provided for the IP trace analysis under the cloud environment. The cloud based traceback services are build with privilege control models. The Framework for Authentication in Cloud-based IP Traceback (FACT) supports the IP traceback operations with security. The FACT scheme performs the user authentication with temporal token details. The traffic flow details and temporal tokens are combined and transferred to the end host. The IP header marking is restricted with header size parameters.

The Enhanced FACT model is build with cloud resource based data origin discovery and security features. The Framework for Authentication in Cloud-based IP Traceback (FACT) scheme is upgraded with optimal marking method. The resource provisioning level based incentive assignment model is adapted in the traceback process. The computational overhead is reduced I the temporal token integrity verification process. The cloud resources and traceback data values are protected under the Enhanced Framework for Cloud based IP Traceback (EFACT) scheme. The Distributed Denial of Service (DDoS) attacks are controlled with service request interval and frequency values.

Index Terms: IP traceback, Cloud resources, Temporal tokens, Attack discovery and Internet Service Provider (ISP)

1. Introduction

Source authentication and path validation are useful primitives to help mitigate various network-based attacks, such as DDoS, address spoofing and flow redirection attacks. Path validation, in particular, provides a way to enforce path compliance according to the policies of ISPs, enterprises and datacenters. Endhosts and ISPs desire to validate service level agreement compliance regarding data delivery in the network: Did the packet truly originate from the claimed client? Did the client select a path that complies with the service provider's policy? Did the packet indeed travel through the path selected by the client?

The current Internet provides almost no means for source authentication and path validation by routers or endhosts, opening up numerous attack surfaces. For example, a malicious ISP may forward a packet on an inferior path while claiming to its client that it forwarded the packet on the premium path. Alternatively, a malicious router may inject packets with a spoofed source address to incriminate a victim source node into having sent an excessive number of packets. A malicious router may simply alter the contents of received packets as well. The inability to detect such attacks near the point of deviation wastes downstream resources. Furthermore, an adversary may exploit the routing protocol to divert traffic to traverse a point of eavesdropping it controls—a serious issue in particular for sensitive information.

End-to-end encryption and authentication mechanisms, such as TLS, do not solve any of the above issues, since they are agnostic to which path the packet takes. A stronger approach is needed, which enables routers and destinations to perform source authentication and path validation. As we discuss in the related work, existing solutions either require extensive overhead, or only partially address

fundamental problems, affecting both feasibility and practicality in the existing network. For example, ICING addresses both source authentication and path validation, but it requires each intermediate router on a path to store and look up keys shared with other routers; ICING requires 42 bytes per verifying router in the packet header. Furthermore, ICING requires each router to calculate a Message Authentication Code (MACs) for all other routers on the path. In contrast, our protocol does not require any perclient state on routers; it requires only 16 bytes per hop and only a single MAC and a single PRF computation per router irrespectively of the path length. Moreover, one of our protocol instantiations prevents against coward attacks, where an adversary only attacks when it knows that the attack will not be detected. Our protocol, offers reduced security in the case of a malicious sender colluding with a malicious router on the path. Since in the common case, sender and receiver trust each other, the performance gain of O(1) MAC operation per router instead of O(n) is worth the tradeoff.

The Dynamically Recreatable Key (DRKey) protocols enable routers to (re-)create symmetric keys shared with the endhosts on the fly. The stateless operation of DRKey on routers prevents state exhaustion DoS attacks and simplifies router architecture. We further enrich DRKey with a new notion called retroactive key setup that provides the following desirable properties: (1) in contrast to previous protocols, source and destination can start the communication without needing to wait for the expensive key setup to complete, providing efficiency; (2) if misbehavior is suspected, endhosts set up keys retroactively to verify previous packets, defending against coward attacks. Based on the dynamically (re-)creatable keys through DRKey protocols, we present Origin and Path Trace (OPT)—-lightweight, scalable and secure protocols for source authentication and path validation. We introduce an extension called Retroactive-PathTrace that supports the destination to perform path validation with retroactive key setup and to detect coward attackers with small, constant overhead in the packet header. Our OPT protocols enable implementation on SW routers with minimal performance impact.

2. Related Work

The essential goal of DDoS attack is to deny the service of a victim through a large volume of requests, such as sending a large amount of ping requests to the victim, or massive request to the victim for downloading large files. Early DDoS attacks emerged around the year 2000 and well-known web sites, such as CNN, Amazon and Yahoo, have been the targets of hackers. The purpose of early attacks was mainly for fun and curiosity about the technique. However, recently we have witnessed an explosive increase in cyber attacks due to the huge financial or political rewards available to cyber attackers [1].

Current, major DDoS attacks are carried out by Botnets. Hackers scan the whole Internet for vulnerable computers, and then compromise them as bots. As a result, an overlay network (botnet) of compromised computers is established, and controlled by botnet masters to commit malicious activities, such as DDoS attacks or information phishing [8]. A DDoS attack can be carried out in various forms, such as flooding packets or synchronization attacks. A recent book by Yu overviews various aspects of DDoS attack and defence in cyberspace [10].

The current dominant traceback mechanism is packet marking, which includes two categories: Probabilistic Packet Marking (PPM) and Deterministic Packet Marking (DPM). In the IPv4 packet head, there are some unused bits, which are usually 16, 17, 19 or 24 bits for different underlay protocols [6]. Network operators can embed special marks or IDs in these available space for traceback purpose. Besides the packet marking mechanism, there are also other mechanisms, such as network traffic based traceback [7]. As they are not directly related with this paper, we do not discuss about them here.

The PPM strategy was firstly proposed and then further improved by researchers, such as in [3]. The basic idea of the PPM scheme is that at the network operator controlled domain, where the victim locates, special marks are injected into the available packet space for incoming packets with a probability at all routers of their domain. At the victim end, we can establish an attack tree based on the received marked packets, and identify the attack sources based on the attack tree.

In order to establish a reliable attack tree, we have to accumulate a large number of marked packets, which causes a challenge on storage and computing power at the victim end. Moreover, the PPM scheme can only trace to the nodes within its domain, which are usually far away from the attacking bots. Different from the PPM method, the DPM scheme deploys a deterministic method and tries to mark packets at routers that are the closest to attack sources (ideally, at the router of the LAN where bots stay). This scheme was firstly proposed by Belenky and Ansari. They noticed that the PPM mechanism can only solve large flooding attacks, and it was not applicable for attacks consisted of a small number of packets. Therefore, they proposed a deterministic packet marking method for IP traceback. The basic idea was that at the initial router of an information source, the router embedded its IP address into the packet by chopping the router's IP into two segments with 17 bits each IP address. As a result, the victim can trace which router the packets came from. Scalability is always critical metric of the DPM schemes. Jin and Yang [9] improved the ID coding of the deterministic packet marking scheme

using redundant decomposition of the initial router IP address. For an IP address, they divided them into three redundant segments, 0 13 bits, 9-22 bits, and 18-31 bits, and then five different hash functions were applied on the three segments to create five results. The resulting eight segments are recorded in the outgoing packets randomly. The victim could reassemble the source router IP using the packets it had received. Furthermore, Xiang, Zhou and Guo [6] noticed the scalability disadvantage of the original DPM scheme, and proposed a flexible deterministic packet marking (FDPM) method to traceback attack sources. They deployed a flexible mark length strategy to match different network environments, and the marking length varied from 16 bits, 19 bits to 24 bits depending on the underneath network protocols. Moreover, they also designed a flexible flow-based marking scheme to adaptively change the marking rate according to the workload of a participating router in the scheme. The FDPM significantly improved the maximum number of traceable sources. For example, for the FDPM-19 and FDPM-24 schemes, they can trace to 8,192 and 262,144 sources, respectively. While the original DPM scheme can only trace to 2,048 sources.

Another interesting traceback method is watermark based strategy. Wang et al. [4] proposed to modulate watermarks into the time interval of a sequence of IP packets at the source side. On the other hand, the receiver can extract the watermark, and further identify the source of the packets. Jia et al. [2] proposed a simple single flow-based scheme to detect the existence of these kind of watermarks in the flow of anonymous communication systems [11]. Moore et al. employed a network telescope technology to observe DDoS activities at a given part of the Internet, 1 256 of the whole IPv4 address space. They collected DDoS attack data for a three year period. Based on their 22 data traces, they found that the average attack event frequency is 24.5/hour. If we extend the observation to the whole Internet, then the average attack even frequency is around 6,272 (24:5 _ 256) per hour. At the same time, they found that the attack durations were relatively short: 60% of attacks were less than 10 minutes, and 80% were less than 30 minutes. Among all attacks, the highest probability of durations were five minutes (10.8% of attacks) and 10 minutes (9.7% of attacks). We will use these key statistics for our experimental part.

In order to estimate the attack power of a DDoS attack, we need to know the size of botnets. However, the research in this part is not that active as attack data is sensitive and hard to obtain from industry. Some reports that the footprint of the Torpig botnet is 182,800, and the median and average size of the Torpigs live population is 49,272 and 48,532, respectively [8]. Our recent collected data set of Conficker indicates that the size of a botnet could be as large as 2,201,183. As we know, bots are compromised computers, due to various reasons (e.g., system reinstallation, power off, anti-virus patching, and so on), the number of active bots of a given botnet is actually far less than their size or footprint. Rajab et al. [5] found that that the number of active bots for a given botnet is usually at the hundreds or a few thousands level. We will also use this information in our experimental part as well.

3. Cloud-Based IP Traceback

IP traceback is an effective solution to identify the sources of packets as well as the paths taken by the packets. It is mainly motivated by the need to trace back network intruders or attackers with spoofed IP addresses, for attribution as well as attack defense and mitigation. For example, traceback is useful in defending against Internet DDoS attacks. It also assists in mitigating attack effects; DoS attacks, for instance, can be mitigated if they are first detected, then traced back to their origins and finally blocked at entry points. In addition, IP traceback can be used for a wide range of

While many different IP traceback approaches have been applied, none of them has achieved universal acceptance or practical deployment. The risk of leaking network topology information ranks as the major challenge in hindering the acceptance of traceback techniques. ISPs (Internet Service Providers) are normally reluctant to allow any external party to gain visibility into their internal structure, since such exposure not only leaks sensitive information to their competitors, but also makes their networks vulnerable to attacks. For example, an adversary may misuse traceback services to reconstruct an ISP's network topology. As a result, ISPs will not wish to participate if the deployment of traceback could leak any sensitive information. Incremental deployability is another important factor for a viable IP traceback solution; it is unrealistic to expect all ISPs to deploy IP traceback services in their networks at the same time. Unfortunately, existing IP traceback mechanisms are inadequate in providing guarantees on privacy and support for incremental deployment. Besides technical shortcomings, economic inefficiency, such as lack of financial incentive for ISPs, also hinders the practical deployment of existing traceback solutions.

The advent of cloud services, offer a new appealing option to support IP traceback service over the Internet. It provides an opportunity to design a traceback system that is incrementally deployable. Cloud storage also increases the feasibility of logging traffic digests for forensic traceback. With a proper access control mechanism, cloud-based traceback can alleviate ISP's privacy concerns of disclosing its internal network topology. In addition, the pay-per-use nature of cloud service provides incentives to

encourage ISPs to deploy traceback service in their networks. Consequently, migrating traditional traceback solutions to cloud becomes more of a natural choice.

The cloud-based traceback architecture exploits increasingly available cloud infrastructures for logging traffic digests, in order to implement forensic traceback. Such cloud-based traceback simplifies the traceback processing and makes traceback service more accessible. It not only possesses privacy-preserving and incremental deployment properties, but also increases robustness against attacks and presents high financial motivation. Yet, regulating access to cloud-based traceback service becomes an important problem. The cloud based IP traceback, named FACT, which enhances traditional authentication protocols such as the password-based scheme in cloud based traceback. The key idea is to embed temporal access tokens in traffic flows and then deliver them to end-hosts in an efficient manner. The method not only ensures that the user requesting for traceback service is an actual recipient of the packets to be traced, but also adapts to the limited marking space in IP header.

4. Issues on Cloud based IP traceback schemes

Cloud based traceback systems are build with access control policies. Framework for Authentication in Cloud-based IP Traceback (FACT) is adapted for authenticating traceback service queries**.** FACT is a temporal token based authentication framework used in cloud environment. FACT embeds temporal access tokens in traffic flows and then delivers them to end-hosts in an efficient manner. The following issues are identified from the current cloud based IP traceback schemes.

- Complex token delivery process
- The marking scheme is not optimized
- Incentive management process is not supported
- Service request based attacks are not efficiently handled

5. Cloud based Data Origin Detection with Security

The data communication over the web and the Intranet environment are performed through the Internet Protocol (IP) support. All the data transmission operations are initiated with the source and destination addresses. The actual source discovery is the main requirement in the network diagnosis applications. IP traceback methods are applied to discover the source and traversed paths of packets. All the network communication operations are registered under the logs maintained under the Internet Service Providers. The ISP logs are analyzed for the source discovery process. The traverse path can also identify using the logs. The log data analysis is a complex task that requires huge computational power and memory. The topology structure of the Internet Service Provider can be accessed by the users. The topology data leakage may increases the security violations under the ISP environment.

The cloud environment provides resources to process the log data values. Log data values are maintained under the cloud storage environment. All the ISP log analysis operations are carried out using the cloud resources. Cloud based traceback architecture is build with traceback services deployed in Internet Service Providers (ISPs). Framework for Authentication in Cloud-based IP Traceback (FACT) is used to handle IP traceback queries. The traceback coordinator and traceback server are deployed under the cloud environment. All the ser traceback query values are collected by the traceback coordinator. The traceback coordinator reroutes the query to the relevant traceback server. The traceback server analyzes the traceback query and discovers the source and its traverse path details. The data processing resources are provided with reference to the temporal token issed b the traceback coordinator. The traceback server verifies the temporal tokens and provides the resources and log data to the users.

The FACT scheme is enhanced with optimal marking scheme, attack control and incentive management mechanisms. The Incentive based Framework for Authentication of Cloud based IP Traceback services (IFACT) scheme is build to support the cloud based traceback operations with higher security and attack control features. The Framework for Authentication in Cloud-based IP Traceback (FACT) is enhanced with optimal marking schemes. Incentive estimation for ISPs is integrated with the system. The ISP protection is improved to handle service request based attacks. Token digest management overhead is controlled in the system.

The secured source discovery for Internet data communication system is build with cloud resources. The Internet Service Provider (ISP) logs are analyzed to discover the source address the for the data communication process. The log files are transferred and analyzed under the cloud resources. The temporal tokens are issued to the users to execute the traceback queries under the cloud environment. The authentication operations are carried out using the temporal tokens. The token verification operations are used confirm the user access to the ISP logs and cloud resources. The incentive model is used to improve the traceback services. The marking problem is solved with service based attack control mechanism. The system has the following advantages. The system protects the topology information of the Internet Service Providers. High scalable and incremental traceback model is proposed. Efficient marking space management mechanism is employed in the system. Privacy and security ensured attack resistant traceback model.

The internet attack discovery operations are carried out using the IP traceback method. The cloud

resources are used to support IP traceback operations. Cloud storage and computational resources are provided for the IP traceback operations. Authentication is provided with temporal tokens. Resource usage time limit is verified with the tokens. The cloud based IP traceback system is divided into five major modules. They are Traceback Coordinator, Traceback Servers, Authentication Process, Attack Discovery and Traceback Process.

The traceback coordinator manages the resources and traceback requests. Traceback servers are placed to interact with the ISPs. User verification process is carried out under the authentication process. Service request based attacks are handled in attack discovery process. Traceback process delivers the traceback results to the users.

5.1. Traceback Coordinator

The traceback coordinator is central authority in the cloud based source discovery application. All the user traceback query values are collected and updated by the traceback coordinator. The traceback coordinator manages the traceback servers with its resources and log data. The query rerouting process redirects the user query values to the relevant traceback server. The load levels are managed with the query redirection procedures. The traceback coordinator issues the temporal tokes to the users for the authentication process. The temporal tokens are used for the authentication process with time boundaries. The temporal token indicates the user identification, requested query value and time limit for the resources. All the resources are provided with reference to the temporal token details. The Internet Service Providers log data are maintained under the cloud environment under the control of the traceback coordinator.

5.2. Traceback Servers

The cloud based source discovery system is build with one traceback coordinator and a set of traceback servers. The traceback servers are used to provide the resources for the source discovery process. The log data values are analyzed under the traceback servers. The traceback server receives the traceback query from the traceback coordinator. The traceback server sends the query response to the relevant user. The log data analysis and source discovery operations are processed within the time limit mentioned in the temporal tokens. The traceback query values are fetched with temporal tokens. The user authentication is varied out with reference to the temporal tokens. The log data access level can also provided in the temporal token structure. The source discovery and path detection results are transmitted to the users from the traceback server. The incentive values are also assigned by the coordinator to the traceback servers and hosts that provides the log data values.

5.3. Authentication Process

The authentication process is carried out under the traceback server environment. The user authentication and access level authentication are carried out for each traceback request. The user authentication verifies the user for the traceback query value. The time limits are verified for the access level authentication. Resources and data limits are verified in the access level authentication for the users. The temporal tokens information are verified in the authentication process. The temporal tokens are protected with the temporal token digest information. All the resources are provided with reference to the authentication results.

5.4. Attack Discovery

The attack discovery process is used to detect and control the service request based attacks. The Distributed Denial of Service (DDoS) attacks are detected in the system. Service request interval and its frequency levels are compared in the attack discovery process. The source address and request similarity are also verified in the attack discovery process. The attack discovery model protects the cloud resources and log data from the malicious users.

5.5. Traceback Process

The traceback process is build to discover the source address for the traceback query values. The traceback server collects the traceback query and parses the query to fetch the request information. The log data analyzed with reference to the query value. The marked data values are compared to fetch the traverse path values. Spoofed data packets are detected in the traceback process. The source discovery results are passed to the users through the traceback server. The traceback process is carried out with the time limit given under the temporal token issued with the traceback query by the traceback coordinator. The traceback query results are updated into the log files.

6. Conclusion and Future Work

The cloud authentication based IP traceback discovery process is developed to utilize the cloud resources for the IP traceback operations. The temporal authentication scheme is used to verify the user requests. The topology protection is provided with the temporal authentication process. Disclosing ISP's internal network topologies, poor incremental deployment and lack of incentives for ISPs parameters are used. The FACT scheme is enhanced to support incentive management model. Attack control schemes are integrated with the system. Optimal marking models are used to manage IP headers. The (FACT) is improved with incentive scheme and attack resistant models. The IP traceback operations can be performed with privacy preserved query submission models. The traceback data values can be maintained and processed in encrypted cloud storage environment.

REFERENCES

[1] T. Peng, C. Leckie, and K. Ramamohanarao, "Survey of networkbased defense mechanisms countering the dos and ddos problems," ACM Computing Survey, vol. 39, no. 1, 2007.

[2] W. Jia, F. P. Tso, Z. Ling, X. Fu, D. Xuan, and W. Yu, "Blind detection of spread spectrum flow watermarks," in INFOCOM 2009. 28th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 19-25 April 2009,

[3] M. T. Goodrich, "Probabilistic packet marking for large-scale ip traceback," IEEE/ACM Trans. Netw., vol. 16, no. 1, pp. 15–24, 2008.

[4] X. Wang, S. Chen, and S. Jajodia, "Network flow watermarking attack on low-latency anonymous communication systems," in 2007 IEEE Symposium on Security and Privacy (S&P 2007), 20-23 May 2007, Oakland, California, USA, 2007, pp. 116–130.

[5] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "My botnet is bigger than yours (maybe, better than yours): why size estimates remain challenging," in HotBots'07: Proceedings of the FIRST conference on Hot Topics in Understanding Botnets, 2007.

[6] Y. Xiang, W. Zhou, and M. Guo, "Flexible deterministic packet marking: An ip traceback system to find the real source of attacks," IEEE Trans. Parallel Distrib. Syst., vol. 20, no. 4, pp. 567– 580, 2009.

[7] S. Yu, W. Zhou, R. Doss, and W. Jia, "Traceback of ddos attacks using entropy variations," IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 3, pp. 412–425, 2011.

[8] B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski, R. Kemmerer, C. Kruegel, and G. Vigna, "Your botnet is my botnet: Analysis of a botnet takeover," in Proceedings of the 2009

[9] G. Jin and J. Yang, "Deterministic packet marking based on redundant decomposition for ip traceback," IEEE Communications Letters, vol. 10, no. 3, pp. 204–206, 2006.

[10] S. Yu, Distributed Denial of Service Attack and Defence. Springer, 2014.

[11] Shui Yu, Wanlei Zhou, Song Guo and Minyi Guoy, "A Feasible IP Traceback Framework through Dynamic Deterministic Packet Marking", IEEE Transactions on Computers, Volume 65, Issue 5 , May 2016.