BlindBox - Deep Packet Inspection Over Encrypted Traffic

Nilesh Parshotam Rijhwani
3771253
nilesh.rijhwani@stud.uni-heidelberg.de

***Abstract*—This seminar explores the challenges of performing Deep Packet Inspection (DPI) on encrypted network traffic, a critical issue in modern cybersecurity. It introduces BlindBox, a novel system that enables DPI directly on encrypted traffic without decryption, thereby preserving user privacy. The seminar delves into the technical details of BlindBox, including its architecture, encryption schemes, and protocols. It also presents an evaluation of BlindBox's functionality and performance, demonstrating its effectiveness in supporting real-world DPI applications. Furthermore, the seminar discusses potential challenges and future directions for BlindBox, highlighting its potential to transform the landscape of network security and privacy.**

I. Introduction

The increasing prevalence of encryption in modern communication, while crucial for safeguarding user privacy, poses a significant challenge to network security tools that rely on Deep Packet Inspection (DPI). This seminar explores this critical conflict and introduces BlindBox, a groundbreaking system that enables DPI directly on encrypted traffic without compromising privacy. We will delve into the technical intricacies of BlindBox, evaluate its performance in real-world scenarios, and discuss its potential to revolutionize the landscape of network security and privacy.

II. Background

In today's digital landscape, the use of encryption protocols like HTTPS has become ubiquitous, safeguarding user privacy by shielding sensitive data from prying eyes. However, this widespread adoption of encryption presents a significant challenge to network security tools that rely on Deep Packet Inspection (DPI) to identify and mitigate threats. DPI, a technique employed by network middleboxes, involves inspecting the payload of network packets, enabling functions such as intrusion detection, data loss prevention, and parental controls. The inherent conflict between the need for privacy through encryption and the necessity for security through DPI creates a complex dilemma.

Current solutions that attempt to bridge this gap often resort to compromising user privacy. Man-in-the-middle (MitM) attacks, where the middlebox decrypts traffic by impersonating the intended recipient, violate the fundamental principles of end-to-end encryption. This approach raises serious concerns about surveillance, data breaches, and unauthorized access to sensitive information, eroding user trust in network security tools.

BlindBox emerges as a promising solution to this predicament. By enabling DPI directly on encrypted traffic without the need for decryption, BlindBox offers a way to maintain both privacy and security. This innovative system leverages a combination of new protocols and encryption schemes to achieve this delicate balance, providing a potential paradigm shift in how we approach network security in an increasingly privacy-conscious world.

III. BlindBox in Detail: Protocols and Privacy Models

BlindBox introduces a novel approach to DPI, operating directly on encrypted traffic without decryption. It achieves this through a combination of specialized protocols and encryption schemes, offering two distinct privacy models:

1) Exact Match Privacy: In this model, the middlebox only learns the positions of exact matches to predefined attack keywords within the encrypted traffic. It gains no knowledge about the content surrounding these keywords or any other part of the traffic. This model is suitable for applications like watermarking, parental filtering, and certain types of intrusion detection that rely on specific keyword matching.
2) Probable Cause Privacy: This model provides even stronger privacy guarantees. The middlebox can only decrypt and inspect a flow if it contains a known attack keyword. If no suspicious keywords are detected, the traffic remains entirely private and inaccessible to the middlebox. This model is ideal for applications that require more comprehensive DPI, such as full-fledged intrusion detection systems, while still upholding user privacy unless there's a valid reason for suspicion.

BlindBox's ability to support these privacy models hinges on its core mechanisms, which we will explore in the following sections. These mechanisms include tokenization, DPIEnc encryption, the BlindBox Detect protocol, obfuscated rule encryption, and probable cause decryption. By understanding these components, we can gain a deeper appreciation for how BlindBox achieves its remarkable balance between network security and user privacy.

IV. Unveiling the Encryption: DPIEnc and BlindBox Detect

BlindBox's core functionality relies on two innovative mechanisms: DPIEnc, a searchable encryption scheme, and the

BlindBox Detect protocol. These mechanisms work in tandem to enable efficient and privacy-preserving DPI on encrypted traffic.

DPIEnc Encryption:

DPIEnc is a specialized encryption scheme designed for BlindBox. It encrypts each token (a fixed-length substring of the data) using a combination of a secret key, a random salt value, and a large prime number for modulo reduction. This encryption ensures that the middlebox cannot directly decipher the token's content but can still perform efficient matching operations.

BlindBox Detect Protocol:

The BlindBox Detect protocol facilitates the middlebox's search for matches between encrypted tokens and encrypted rules (signatures of known attacks). It involves precomputation by the middlebox, where it generates encrypted rule values for all possible salt values. These values are organized into a search tree for fast lookup. When the middlebox receives encrypted tokens, it compares them against the search tree to identify potential matches. The sender maintains a counter table to ensure unique salts for each token, preventing information leakage.

Together, DPIEnc and BlindBox Detect enable the middlebox to efficiently search for specific keywords or patterns within the encrypted traffic without revealing the actual content of the data. This capability forms the foundation for BlindBox's privacy-preserving DPI functionality.

V. The Art of Concealment: Obfuscated Rule Encryption

BlindBox employs Obfuscated Rule Encryption to ensure that the middlebox can obtain encrypted rules without learning the actual rules or the endpoints' secret key. This technique leverages two cryptographic primitives:

1) Yao's Garbled Circuits: These circuits enable secure two-party computation, allowing the middlebox and endpoints to jointly compute encrypted rule values without revealing their inputs. The rule generator constructs garbled circuits representing the rule logic, and the endpoints evaluate these circuits using their secret key, producing encrypted rule values that the middlebox can use for matching.
2) Oblivious Transfer: This protocol allows the middlebox to selectively obtain encrypted rule values from the endpoints without revealing which rules it is interested in. The endpoints provide a set of encrypted rule values, and the middlebox chooses which ones to receive based on its needs, ensuring that the endpoints remain oblivious to the middlebox's specific rule selection.

Obfuscated Rule Encryption adds a crucial layer of protection to BlindBox, safeguarding the confidentiality of the rules used for DPI while still enabling the middlebox to perform its function effectively.

VI. Selective Transparency: Probable Cause Decryption

BlindBox's Probable Cause Privacy model introduces a unique mechanism called Probable Cause Decryption, allowing the middlebox to decrypt a flow only when a suspicious keyword is detected. This selective decryption capability ensures that user traffic remains private unless there's a legitimate reason for inspection, striking a balance between security and privacy.

How it Works:

1) Embedded SSL Key: The sender embeds the SSL session key within the encrypted tokens in a way that can only be recovered if a match with a suspicious keyword occurs.
2) Keyword Match: When the middlebox detects a match, it uses the information from the matching token to reconstruct the embedded SSL key.
3) Selective Decryption: With the recovered SSL key, the middlebox can decrypt the specific flow containing the suspicious keyword, enabling further analysis using traditional DPI tools.
4) Privacy Preservation: If no match is found, the SSL key remains unrecoverable, and the traffic remains encrypted and private.

Probable Cause Decryption empowers the middlebox to investigate potential threats without compromising the privacy of benign traffic. This mechanism is particularly crucial for applications like intrusion detection, where a deeper inspection might be necessary to confirm the presence of an attack.

VII. Implementation

BlindBox's implementation involves both the client/server side and the middlebox, requiring modifications to existing protocols and the development of specialized libraries.

System Implementation:

1) BlindBox HTTPS: A C library that handles the transmission of encrypted traffic and tokens. It integrates with the GnuTLS library to extract session keys for probable cause decryption.
2) Click-based Middlebox: A multi-threaded implementation using the Click modular router framework and DPDK for high-performance packet processing. It includes detection threads for traffic inspection and garble threads for handling garbled circuits.

BlindBox HTTPS Protocol:

1) Socket Management: Opens three separate sockets for standard SSL traffic, searchable encrypted token transmission, and garbled circuit exchange.
2) Protocol Modifications: Modifies the GnuTLS library to extract session keys for probable cause decryption (Protocol III).

This implementation demonstrates BlindBox's practicality and its ability to integrate with existing network infrastructure, paving the way for its potential deployment in real-world scenarios.

VIII. Evaluation

BlindBox's evaluation focuses on assessing its functionality in supporting real-world DPI applications and measuring its performance impact on both the client and network sides.

Functionality Evaluation:

The evaluation examines BlindBox's ability to handle various DPI rule sets, including those used for document watermarking, parental filtering, and intrusion detection. It analyzes the coverage of different BlindBox protocols (I, II, and III) for these applications. The findings indicate that BlindBox can effectively implement the functionality required for most DPI applications, with the choice of protocol depending on the complexity of the rules and desired privacy level.

Performance Evaluation:

1) Client-Side: Micro-benchmarks and realistic scenarios are used to measure the overhead introduced by BlindBox on the client side. The evaluation considers encryption time, handshake time, page load time, and bandwidth overhead. While BlindBox adds some overhead compared to standard HTTPS, it remains efficient enough for practical use, especially with hardware acceleration. The main overhead is the initial handshake for obfuscated rule encryption, which scales with the number of rules.
2) Middlebox: The evaluation assesses BlindBox's throughput and compares it with alternative approaches like searchable encryption and functional encryption. BlindBox demonstrates high throughput, competitive with standard IDS implementations, due to efficient token matching and offloading complex processing. It significantly outperforms the strawman approaches in terms of detection time, highlighting its specialized design for DPI over encrypted traffic.

IX. Challenges: Adoption and Privacy Considerations

While BlindBox presents a promising solution, its widespread adoption faces several challenges.

1) ISP Adoption: Internet Service Providers (ISPs) might hesitate to embrace BlindBox due to potential conflicts with their existing business models, which often rely on collecting and analyzing user data. Regulatory intervention or changes in incentives, such as offering BlindBox as a premium service, might be necessary to encourage ISP adoption.
2) Client Adoption: BlindBox necessitates modifications to the standard HTTPS protocol, potentially hindering its widespread adoption. Users might need to install new software or browser extensions to utilize BlindBox HTTPS. Educating users about the benefits of BlindBox and addressing concerns about compatibility and usability will be crucial for its acceptance.

Privacy considerations also play a vital role in BlindBox's deployment.

1) Rule Design: The choice of rules significantly impacts privacy. Overly broad or generic rules could lead to unintended information leakage. Careful rule design is essential to strike a balance between security needs and privacy preservation.
2) Tokenization Impact: The tokenization strategy influences the amount of information revealed to the middlebox. Window-based tokenization offers comprehensive coverage but might disclose more information than delimiter-based tokenization. The choice of tokenization strategy should consider the trade-off between detection accuracy and privacy.

Addressing these challenges and privacy concerns will be crucial for BlindBox to realize its full potential in transforming the landscape of network security and privacy.

X. Beyond DPI: Expanding BlindBox's Horizons

While BlindBox's primary focus lies in enabling DPI on encrypted traffic, its underlying principles and techniques hold the potential for broader applications in the realm of network security and privacy.

One promising avenue for future exploration is extending BlindBox's capabilities to other types of middleboxes beyond those performing DPI. This could include network appliances like caches, protocol accelerators, and compression engines, which currently face similar challenges in operating effectively on encrypted traffic. By adapting BlindBox's core mechanisms, such as tokenization and selective decryption, it might be possible to enable these middleboxes to function securely and privately without compromising the benefits of encryption.

Another area for future work lies in enhancing BlindBox's performance, particularly for short-lived connections and large rulesets. The current implementation incurs some overhead during the initial handshake for obfuscated rule encryption, which can be significant for scenarios with numerous rules. Optimizing this process or exploring alternative approaches could make BlindBox more suitable for a wider range of network environments and applications.

Furthermore, strengthening BlindBox's privacy guarantees is an ongoing endeavor. While the existing privacy models offer substantial protection, there's always room for improvement. Exploring techniques like all-or-nothing rule matching, where the middlebox only learns if an entire rule matches rather than individual keywords, could further minimize information leakage. Additionally, conducting a rigorous formal privacy analysis under various threat models could provide valuable insights into potential vulnerabilities and guide the development of even stronger privacy protections.

XI. Conclusion

The seminar delves into the innovative BlindBox system, a solution to the challenge of performing Deep Packet Inspection (DPI) on encrypted network traffic without compromising user privacy. It explores BlindBox's technical underpinnings, evaluates its performance, and discusses its potential to reshape network security and privacy. BlindBox offers a promising path toward a future where robust security measures can coexist with strong privacy protections, ensuring a safer and more private online experience for all.

References

[1] J. Sherry, C. Lan, R. A. Popa, and S. Ratnasamy, ”Blindbox: Deep packet inspection over encrypted traffic,” in *Proceedings of the 2015 ACM SIGCOMM Conference*, 2015, pp. 213-226.

[2] Dyer, K. P., Coull, S. E., Ristenpart, T., Shrimpton, T. (2012). Peek-a-boo, I still see you: Why efficient traffic analysis countermeasures fail. In 2012 IEEE Symposium on Security and Privacy (pp. 332-346). IEEE.

[3] Bellovin, S. M., Goldberg, I. (2018). Encrypted content inspection: Threats and responses. IEEE Security Privacy, 16(1), 70-75.

[4] Springall, D., Kitcat, J., Danezis, G. (2017). Privacy preserving deep packet inspection. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 2101-2118).

[5] Van Oorschot, P. C. (2016). Revisiting man-in-the-middle attacks on SSL/TLS. IEEE Security Privacy, 14(6), 60-67.