Deniz Aytemiz

CS5590-System and Software Security: Paper Summaries for Feb 21st

Privacy-Preserving Detection of Sensitive Data Exposure:

The solution presented in this paper tackles the problem of inadvertent data leaks occurring on supervised network channels with a different approach. In their model, the leak detection processed by DLD provider is considered as semi-honest, meaning that, it should be able to perform deep package inspection but without inferring the unencrypted/plaintext sensitive data coming from source. Hence they use a modified content based DLPD approach. This approach involves disclosing only a portion of sensitive data and sharing fuzzy fingerprints of the data which hides sensitive data among unsentive data so that the DLD provider cannot pinpoint. The DLD provider returns some alerts, in other words, possible data-leak cases. Later, during the post process, the data owner distinguishes between the false positives and actual data leaks. Examining the evaluation reports, this solution works very well (quantitatively) in some scenarios though there are some areas that need improvement. For instance, the variable 'fuzzy length' introduces some trade off between speed of post process and better privacy protection against DLD. In addition it can be vulnerable against leaks in case of heavily modified data since the fingerprints end up very different from unmodified data's fingerprints.[1]

Enterprise data breach: causes, challenges, prevention, and future directions:

In the era of big data, enterprises face severely damaging cyber attacks both in terms of reputation and financially. This paper reviews data leak threats, the techniques used in DLPD recently in addition to promising developments and current challenges. Data leak incidents can be categorized as external and internal depending on who causes the breach. Internal data leaks are generally more difficult to detect since the insider has the authority of access to data or facilities and may be aware of bypassing countermeasures. External data leaks on the other hand are more trivial to overcome with current DLPD techniques as discussed in the case of Target breach in 2016. DLPD techniques can be categorized as basic security measures(firewall, anti-virus softwares etc) and designated DPLD techniques which are being used at multiple different points in the enterprise environment. Designated DLPD techniques can also be categorized as content and context based depending on the technology employed in it. Content based

approach is the one discussed in the paper named ‘Privacy-Preserving Detection of Sensitive Data Exposure’. Content based approach usually achieves more accuracy then context based. However it has its own shortcomings as being prone to bypassing internal attack or data obfuscation and has scalability issues since large amounts of data needs to be processed. In the context based approach there are methods built on building a model of usual insider behavior and detecting anomalies in the usual pattern in order to detect malicious insider attempts. These methods rely on data mining and machine learning mostly and they bring the challenge of insufficient data training set. There are also watermarking and honeypot techniques but they also have vulnerabilities especially against internal attacks hence the malicious insider is an ongoing problem researchers are focusing on. The research areas highlighted in this paper can be summarized as; anomaly detection with deep learning for behavioral analysis, scalable cloud environment that can perform data leak detections with high accuracy and insignificant delays during processing, methods to monitor encrypted channel such that DLPD techniques will not be inapplicable against encrypted data leakage and a common data set for DLPD techniques such that they can be tested, evaluated and compared in a standardized/measurable manner.[2]

Actions Taken by Equifax and Federal Agencies in Response to the 2017 Breach:

This paper summarizes the events regarding the breach of Equifax in 2017. During the breach, much personal identifying data was extracted from Equifax’s databases. This did not only affect Equifax but also the companies acquired Equifax’s services for identity verification such as IRS, USPS and SSA. In March 2017, attackers used scanning softwares in Equifax network to pinpoint a vulnerable point and identified a vulnerability in Equifax’s online portal that subsequently allowed them to gain unauthorized access and run commands. The commands and queries the attacker used were encrypted and kept in small chunks in order to be able to bypass the security countermeasures of Equifax and since there was an expired digital certificate, encrypted traffic was uninspected. Hence almost for 3 months the breach was undetected causing much sensitive data to be breached and breached from multiple databases since individual databases were not isolated. After breach was discovered by Equifax, an extensive investigation began on the scope of breach and the identity of customers victimized by the attack. The actions of attackers were analyzed by the help of log files. The customers are notified and Equifax offered some free services for the clients to protect themselves from the harmful consequences of the breach. Equifax also acknowledges the customer

agencies that were customers to them and those agencies themselves conducted individual assessments on the breach, notified its customers and increased their security measures. Agencies also reviewed the security of Equifax's data center, modified their contract for potential future breaches, improved their own identity verifying procedure and canceled their contract with Equifax. After the breach Equifax went under extensive security improvements by the help of an external cybersecurity consultant and was required to submit a list of remediation projects being implemented to banking regulators. One can conclude the breach had a significant harm on Equifax, agencies dependent on it and the customers. [3]

Bibliography:

[1] **[Shu 2015]** Xiaokui Shu, Danfeng Yao, and Elisa Bertino. Privacy-Preserving Detection of Sensitive Data Exposure. IEEE Transactions on Information Forensics & Security (TIFS). 10(5). 1092-1103. May 2015.
[2] **[Cheng 2017]** Long Cheng, Fang Liu, and Danfeng Yao. Enterprise Data Breach: Causes, Challenges, Prevention, and Future Directions. WIREs Data Mining and Knowledge Discovery. Wiley. 2017.
[3] **[Equifax Breach Report 2018]** Actions Taken by Equifax and Federal Agencies in Response to the 2017 Breach. Government Accountability Office. 2018. https://www.gao.gov/assets/gao-18-559.pdf