Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Masterarbeit

Exposé zur Masterarbeit - Pseudonymisierung von und Einsatz von Schwellwertschemata für Logeinträge

vorgelegt von

Tom Petersen

geb. am 13. Dezember 1990 in Hannover

Matrikelnummer 3659640

Studiengang Informatik

eingereicht am 10. Juli 2017

Betreuer: Dipl.-Inf. Ephraim Zimmer

Erstgutachter: -

Zweitgutachter: -

1 Einführung

Im Folgenden soll zuerst das Thema der Arbeit motiviert und anschließend auf mögliche Schwerpunkte, die für die Arbeit gesetzt werden können, eingegangen werden.

1.1 Motivation

Liest man von erfolgreichen Angriffen auf Unternehmensnetzwerke, so ist die implizite Annahme von außenstehenden, unternehmensfremden Angreifern weit verbreitet. Doch häufig sind die Angreifer bereits im Netzwerk ansässig. Es handelt sich um (ehemalige) Mitarbeiter oder zumindest Personen mit legitimem Zugriff auf das Netzwerk, wie Geschäftspartnern oder Kunden. Hierbei geht es keineswegs lediglich um Einzelfälle.

In dem *IBM Cyber Security Intelligence Report* von 2015 werden 55% der Angriffe als aus dem internen Netz stammend angegeben [Bra+15]. Zu beachten ist allerdings, dass nicht nur mit Absicht ausgeführte Angriffe hierunter erfasst wurden, sondern auch unbeabsichtigte wie das versehentliche Veröffentlichen schützenswerter Kundendaten.

Auch der Branchenverband bitkom führt in seiner *Spezialstudie Wirtschaftsschutz* aus dem Jahr 2016 nach einer Befragung von über 1000 Unternehmen aus, dass etwa 60% der erfolgten Handlungen aus dem Bereich Datendiebstahl, Industriespionage oder Sabotage durch (ehemalige) Mitarbeiter erfolgten [bit16].

Auch wenn die genauen Zahlen aufgrund von unterschiedlichen Annahmen und der in diesem Bereich nicht zu vernachlässigenden Dunkelziffer[1] mit Vorsicht zu betrachten sind, so geben sie doch Hinweise darauf, dass Angriffe von Innentätern weit verbreitet sind und ein hohes Schadenspotenzial aufweisen. Die Erkennung und Verhinderung solcher Angriffe sollte daher ein wichtiger Teil des IT-Sicherheitskonzepts eines Unternehmens sein.

Zur Erkennung von Angriffen in Netzwerken können SIEM-Systeme eingesetzt werden (siehe Abschnitt 2.1). Diese sind jedoch in erster Linie auf das Erkennen von externen Angriffen ausgelegt und in ihrer derzeitigen Form kaum sinnvoll für das Erkennen von Innentätern zu nutzen.
Hierfür würden zusätzliche Datenquellen und Erkennungslogiken nötig sein. Zusätzlich spielen auch datenschutzrechtliche Bedenken im Bezug auf das Sammeln von großen Datenmengen über Mitarbeiter des eigenen Unternehmens hier eine entscheidende Rolle.

Ein Ansatz, der diese Bedenken ausräumen oder zumindest lindern könnte, ist die Nutzung von Pseudonymen bei der Datenerfassung (siehe Abschnitt 2.2). Anstatt direkt identifizierende Merkmale eines Arbeitnehmers abzuspeichern, werden diese Merkmale durch ein Pseudonym ersetzt. Eine Liste dieser Ersetzungen wird verschlüsselt abgelegt. Im Fall eines Angriffs durch einen Innentäter kann die Liste entschlüsselt werden und relevante Ereignisse de-pseudonymisiert,

1. Insbesondere die Angst vor Imageschäden, die auch in der *Spezialstudie Wirtschaftsschutz* erwähnt wird, könnte ein Grund für das Geheimhalten von Vorfällen sein.

also ihrem ursprünglichen Verursacher wieder zweifelsfrei zugeordnet, werden.
Um die Entschlüsselung nicht einzelnen (möglicherweise bösartig agierenden) Personen zu ermöglichen, können sogenannte Schwellwertschemata eingesetzt werden (siehe Abschnitt 2.3). Durch sie wird die Entschlüsselung erst durch die Kooperation mehrerer Parteien möglich gemacht.

Bei diesem Ansatz muss jedoch auch beachtet werden, dass durch den Einsatz von Pseudonymen die Erkennung von Angriffen erschwert werden könnte. Beispielsweise könnte das Ändern von Pseudonymen in regelmäßigen Zeitintervallen und die dadurch entstehende Nicht-Verkettbarkeit von Ereignissen dafür sorgen, dass längfristig angelegte Angriffe nicht aufgedeckt werden.

1.2 Ziele der Arbeit

In dieser Arbeit soll es darum gehen, prototypisch ein solches Szenario auf Basis eines Open-Source-SIEM-Systems umzusetzen. Hierbei müssen einige Fragen betrachtet werden:

- An welcher Stelle des Systems kann eingegriffen werden, um die erfassten Daten zu verändern, und welche Auswirkungen hat dies?
- Wie erfolgt die angesprochene Pseudonymisierung technisch?
- Welche kryptographischen Schwellwertschemata können genutzt werden? Gibt es bereits quelloffene Implementierungen? Was muss selbst entwickelt werden? Wie erfolgt das Schlüsselmanagement?
- Können neben der Pseudonymisierung noch weitere Funktionen zur Veränderung von Daten sinnvoll sein und wie könnten diese umgesetzt werden?

Gerade die letzte Frage sorgt dafür, dass zusätzliche Anforderungen an den zu entwickelnden Prototypen gestellt werden. Es sollte möglich sein, abhängig von den eingehenden Daten die entsprechend gewünschten Funktionen konfigurieren und den Prototypen in eventuell aufbauenden Arbeiten auch um zusätzliche Funktionen ergänzen zu können.

2 Inhalte

In diesem Kapitel sollen einige theoretische Hintergründe für die für diese Arbeit relevanten Themen dargelegt werden.

2.1 SIEM-Systeme

Der Begriff SIEM (Security Information and Event Management) setzt sich aus SEM (Security Event Management), das sich mit Echtzeitüberwachung und Ereigniskorrelation befasst, sowie SIM (Security Information Management), in dessen Fokus Langzeiterfassung und Analyse von Log-Daten steht, zusammen [NK11]. SIEM-Systeme dienen dazu Daten in Netzwerken zu sammeln, um so einen zentralisierten Überblick über das Netzwerk zu erhalten und Bedrohungen erkennen und verhindern zu können.

Ein SIEM-System sollte unter anderem die folgenden Aufgaben erfüllen:

- Event-Behandlung
- Erkennung von Anomalien auf Netzwerkebene
- Überprüfung der Einhaltung von Richtlinien (Compliance Reporting)
- Bereitstellung von Schnittstellen zur Integration heterogener Systeme im Netzwerk
- Nutzerabhängige Sichten auf sicherheitsrelevante Ereignisse

Details dazu sind [Det+15] zu entnehmen.

Eine besondere Bedeutung kommt hier der Behandlung von sicherheitsrelevanten Ereignissen (Events) zu, die beispielsweise von Intrusion-Detectionen-Systemen oder aus den Log-Daten von Firewalls, Switches,... stammen können. Hier muss ein SIEM-System nach [DRS14] insbesondere drei Aufgaben erfüllen:

- **Extraktion:** Die Daten werden aus Logeinträgen oder empfangenen Systemmeldungen extrahiert.
- **Mapping:** Die extrahierten Daten werden in ein SIEM-spezifisches Format übersetzt, um eine sinnvolle Weiterverarbeitung zu gewährleisten.
- **Aggregation:** Gleichartige Events können in manchen Fällen anschließend zusammengefasst werden, um aussagekräftigere Informationen zu erhalten.

Weiterhin können SIEM-Systeme noch zusätzliche Aufgaben wie Schwachstellenscans oder Netzwerk-Monitoring übernehmen.

Eine quelloffene SIEM-Lösung, die im Rahmen dieser Arbeit genutzt werden wird, ist OSSIM, ein SIEM-System der Firma AlienVault, das auf Basis weiterer quelloffener Lösungen aus dem Netzwerksicherheits-Bereich unter anderem die oben genannten Funktionen bereitstellt[1].

2.2 Pseudonymisierung

Pseudonymisierung beschreibt nach [PK01; PH10] die Benutzung von Pseudonymen zur Identifizierung von Subjekten, wobei ein Pseudonym[2] als Identifikator eines Subjekts ungleich seinem echten Bezeichner definiert wird.

Pseudonymität sagt dabei erst einmal lediglich etwas über die Verwendung eines Verfahrens aus, jedoch nichts über die daraus entstehenden Auswirkungen auf die Identifizierbarkeit eines Subjekts oder die Zurechenbarkeit bestimmter Aktionen. Hierfür spielen weitere Eigenschaften von Pseudonymen wie die folgenden eine Rolle:

- garantierte Eindeutigkeit von Pseudonymen
- Möglichkeit von Pseudonymänderungen
- begrenzt häufige Verwendung von Pseudonymen
- zeitlich begrenzte Verwendung von Pseudonymen
- Art der Pseudonymserstellung

Die Ausprägungen dieser Eigenschaften werden auch im Rahmen dieser Arbeit für das umzusetzende System zu bewerten sein.

2.3 Schwellwertschemata

1976 entwickelte Shamir das erste (k,n)-Schwellwert-Schema: Ein Geheimnis D wird so in n Teile $D_1,\ldots,D_n$ zerlegt, dass durch Kenntnis von mindestens k Teilen das Geheimnis wieder aufgedeckt werden kann, aber jede Kombination aus höchstens $k-1$ Teilen keine Informationen über D liefert. Shamirs Lösung bediente sich der Polynominterpolation auf der Basis modularer Arithmetik [Sha79].
Im selben Jahr veröffentlichte auch Blakley eine Lösung dieses Problems, die auf den Schnittpunkten von Hyperebenen über endlichen Feldern beruht [Bla79].

Das Problem dieser Lösungen bezogen auf den hier behandelten Anwendungsfall ist jedoch, dass das Geheimnis nach erstmaligem Aufdecken bekannt ist. Wünschenswert wäre ein Verfahren, bei dem nur ein entsprechend verschlüsseltes Datum (bspw. der gesuchte Eintrag in einer Pseudonym-Tabelle) aufgedeckt werden kann, ohne dass der kombinierte Schlüssel selbst bekannt wird.

1. AlienVault OSSIM: The World's Most Widely Used Open Source SIEM - https://www.alienvault.com/products/ossim
2. ursprünglich aus dem Griechischen stammend: *pseudonumon* - falsch benannt

In [Des87] wird dieses Problem das erste Mal im Kontext von verschlüsselten Nachrichten an Gruppen betrachtet: Ein Sender möchte eine Nachricht an eine Gruppe von Empfängern senden, die nur in Zusammenarbeit die Nachricht entschlüsseln können sollen. Hier wird auch die zentrale Forderung aufgestellt, den mehrfachen Nachrichtenaustausch zwischen Sender und Empfänger(n) bei der Entschlüsselung (sogenannte Ping-Pong-Protokolle) zu vermeiden.
In [Des93] spricht der Autor bei dieser Klasse von Verfahren von *threshold decryption* und fordert weiterhin, dass praktisch einsatzbare Systeme auch *non-interactive* sein sollten, also bei der Entschlüsselung keinen aufwendigen Datenaustausch zwischen den Teilnehmern notwendig machen.

In [BBH06] werden diese Systeme formalisiert. Ein *Threshold-Public-Key-Encryption*-System besteht aus fünf Schritten:

1. $Setup(n,k,\Lambda)$ liefert ein Tripel (PK,VK,SK), bestehend aus dem öffentlichen Schlüssel PK, einem Verifikationsschlüssel VK und einer n-elementigen Liste aus *Private Key Shares*, von denen jeder Teilnehmer einen *Share* erhält. Λ wird als initialer Sicherheitsparameter bezeichnet.
2. $Encrypt(PK,M)$ liefert die verschlüsselte Nachricht C.
3. $ShareDecrypt(PK,i,SK_i,C)$ liefert ein *Decryption Share* $\mu = (i,\mu^i)$ des i-ten Teilnehmers, das im 5. Schritt zusammen mit weiteren *Shares* zur Entschlüsselung der Nachricht genutzt wird.
4. $ShareVerify(PK,VK,C,\mu)$ überprüft ein *Decryption Share* auf Validität.
5. $Combine(PK,VK,C,\mu_1,\ldots,\mu_k,\ldots)$ verknüpft die *Decryption Shares* von mindestens k Teilnehmern und liefert die Nachricht M zurück.

Anforderungen an diese Schritte sind [BBH06] zu entnehmen.

Ein solches System, das auf dem ElGamal-Algorithmus und damit dem Diskreten-Logarithmus-Problem basiert, veröffentlichten die Autoren in [DF90]. Dieser Ansatz setzt in der *Setup*-Phase auf eine zentrale vertrauenswürdige Stelle zur Erzeugung der Schlüssel und *Shares*. In [Ped91] stellt der Autor basierend auf diesen Ergebnissen ein Verfahren vor, das bei der Schlüsselgenerierung ohne eine vertrauenswürdige Instanz auskommt. Dieses Verfahren wird in [Gen+99] noch einmal verbessert.
Basierend auf dem jetzigen Recherchestand würde sich diese Kombination von Verfahren gut für den angestrebten Anwendungszweck eignen. Konkrete offene Implementierungen wurden jedoch bisher nicht gefunden, so dass möglicherweise eine eigene Implementierung umgesetzt werden muss.

Neben diesem Verfahren gibt es noch weitere Ansätze basierend auf RSA [Des93; Ngu05] oder dem Paillier-Kryptosystem [Pai99; DJ01], die jedoch deutlich komplexer zu sein scheinen.

Literatur

[BBH06] Dan Boneh, Xavier Boyen und Shai Halevi. *Chosen ciphertext secure public key threshold encryption without random oracles*. In: *Topics in Cryptology – CT-RSA 2006*. Hrsg. von David Pointcheval. Springer, 2006, S. 226–243.

[bit16] bitkom. *Spezialstudie Wirtschaftsschutz*. bitkom. 2016. URL: https://www.bitkom.org/Bitkom/Publikationen/Spezialstudie-Wirtschaftsschutz.html (besucht am 24. 06. 2017).

[Bla79] George Robert Blakley. *Safeguarding cryptographic keys*. In: *Proceedings of the AFIPS 1979 National Computer Conference*. Hrsg. von Richar E. Merwin. AFIPS Press, 1979, S. 313–317.

[Bra+15] Nicholas Bradley u. a. *IBM 2015 Cyber Security Intelligence Index*. IBM. 2015. URL: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=SEW03073USEN (besucht am 24. 06. 2017).

[Des87] Yvo Desmedt. *Society and Group Oriented Cryptography: a New Concept*. In: *Conference on the Theory and Application of Cryptographic Techniques 1987*. Hrsg. von Carl Pomerance. Springer, 1987, S. 120–127.

[Des93] Yvo Desmedt. *Threshold cryptosystems*. In: *Advances in Cryptology — AUSCRYPT '92*. Hrsg. von Jennifer Seberry und Yuliang Zheng. Springer, 1993, S. 1–14.

[Det+15] Kai-Oliver Detken u. a. *SIEM approach for a higher level of IT security in enterprise networks*. In: *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. IEEE. 2015, S. 322–327.

[DF90] Yvo Desmedt und Yair Frankel. *Threshold cryptosystems*. In: *Advances in Cryptology - CRYPTO' 89 Proceedings*. Hrsg. von Gilles Brassard. Springer, 1990, S. 307–315.

[DJ01] Ivan Damgard und Mads Jurik. *A generalisation, a simplification and some applications of Paillier's probabilistic public-key system*. In: *Public Key Cryptography – Proceedings of the 4th International Workshop on Practice and Theory in Public Key Cryptosystems*. Hrsg. von Kwangjo Kim. Bd. 1992. Springer, 2001, S. 119–136.

[DRS14] Kai-Oliver Detken, Thomas Rossow und Ralf Steuerwald. *SIEM-Ansätze zur Erhöhung der IT-Sicherheit auf Basis von IF-MAP*. In: *Proceedings of the D A CH Security 2014: Bestandsaufnahme, Konzepte, Anwendungen und Perspektiven*. Hrsg. von Peter Schartner und Peter Lipp. syssec, 2014.

[Gen+99] Rosario Gennaro u. a. *Secure distributed key generation for discrete-log based cryptosystems*. In: *Advances in Cryptology — EUROCRYPT'99*. Hrsg. von Jaques Stern. Springer, 1999, S. 295–310.

[Ngu05] H.L. Nguyen. *RSA Threshold Cryptography*. 2005. URL: https://www.cs.ox.ac.uk/files/269/Thesis.pdf (besucht am 02. 07. 2017).

[NK11] Mark Nicolett und Kelly M Kavanagh. *Magic quadrant for security information and event management*. In: *Gartner Research Note G00212454* (2011).

[Pai99] Pascal Paillier. *Public-key cryptosystems based on composite degree residuosity classes*. In: *Advances in Cryptology — EUROCRYPT'99*. Hrsg. von Jaques Stern. Springer, 1999, S. 223–238.

[Ped91] Torben Pedersen. *A threshold cryptosystem without a trusted party*. In: *Advances in Cryptology — EUROCRYPT'91*. Hrsg. von Donald W. Davies. Springer, 1991, S. 522–526.

[PH10] Andreas Pfitzmann und Marit Hansen. *A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management*. 2010. URL: http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf (besucht am 27. 06. 2017).

[PK01] Andreas Pfitzmann und Marit Köhntopp. *Anonymity, unobservability, and pseudonymity — a proposal for terminology*. In: *Designing privacy enhancing technologies*. Hrsg. von Hannes Federrath. Springer, 2001, S. 1–9.

[Sha79] Adi Shamir. *How to share a secret*. In: *Communications of the ACM* 22.11 (1979), S. 612–613.