

# Contents

<b>1</b>	<b>Introduction to Information Security .....</b>	<b>1</b>
<b>2</b>	<b>Data Leakage.....</b>	<b>5</b>
<b>3</b>	<b>A Taxonomy of Data Leakage Prevention Solutions.....</b>	<b>11</b>
3.1	What to protect? (data-state) .....	11
3.2	Where to protect? (deployment scheme).....	12
3.3	How to protect? (leakage handling approach).....	13
<b>4</b>	<b>Data Leakage Detection/Prevention Solutions .....</b>	<b>17</b>
4.1	A review of commercial DLP solutions .....	17
4.1.1	Market overview.....	17
4.1.2	Technological offerings of market leaders .....	17
4.1.3	Conclusions, remarks, and problems with the state of the art in industrial DLP .....	19
4.2	Academic research in the DLP domain.....	21
4.2.1	Misuse detection in information retrieval (IR) systems .....	24
4.2.2	Misuse detection in databases .....	26
4.2.3	Email leakage protection.....	28
4.2.4	Network/web-based protection .....	30
4.2.5	Encryption and access control.....	31
4.2.6	Hidden data in files.....	33
4.2.7	Honeypots for detecting malicious insiders .....	34
<b>5</b>	<b>Data Leakage/Misuse Scenarios .....</b>	<b>39</b>
5.1	Classification of data leakage/misuse scenarios .....	39
5.1.1	Where did the leakage occur? .....	39
5.1.2	Who caused the leakage? .....	39
5.1.3	What was leaked?.....	40
5.1.4	How was access to the data gained?.....	40
5.1.5	How did the data leak? .....	41
5.2	Description of main data leakage/misuse scenarios .....	41
5.3	Discussion .....	46

<b>6</b>	<b>Privacy, Data Anonymization, and Secure Data Publishing</b>	47
6.1	Introduction to data anonymization	47
6.2	Elementary anonymization operations	48
6.2.1	Generalization	48
6.2.2	Suppression	51
6.2.3	Permutation	51
6.2.4	Perturbation	51
6.3	Privacy models	52
6.3.1	Basic concepts	52
6.3.2	$k$ -Anonymity	52
6.3.3	$L$ -Diversity	54
6.3.4	$K$ -Uncertainty	55
6.3.5	(X,Y)-Privacy	56
6.3.6	(X,Y)-Anonymity	56
6.3.7	(X,Y)-Linkability	57
6.4	Metrics	58
6.4.1	Information metrics	58
6.4.2	Search metrics	59
6.5	Standard anonymization algorithms	60
6.6	Multiple-release publishing	62
6.6.1	Single vs. multiple-release publishing	63
6.6.2	Publishing releases in parallel	63
6.6.3	Publishing releases in sequence	64
6.6.4	Anonymizing sequential releases	64
<b>7</b>	<b>Case studies</b>	69
7.1	Misuse detection in database systems	69
7.1.1	Applying unsupervised context-based analysis	70
7.1.2	Calculating a misusability score for tabular data	74
7.2	Using honeytokens	76
7.3	Email leakage	79
<b>8</b>	<b>Future Trends in Data Leakage</b>	83
	<b>References</b>	87

A Survey of Data Leakage Detection and Prevention  
Solutions

Shabtai, A.; Elovici, Y.; Rokach, L.

2012, VIII, 92 p. 9 illus., Softcover

ISBN: 978-1-4614-2052-1