KNOWLEDGE REPRESENTATION OF SECURITY DESIGN PATTERN LANDSCAPE USING FORMAL CONCEPT ANALYSIS

POONAM S. PONDE*, SHAILAJA C. SHIRWAIKAR, VILAS S. KHARAT

Department of Computer Science, Savitribai Phule Pune University, Pune 411007, India
*Corresponding Author: poonamponde@gmail.com

Abstract

Security design patterns are proven solutions to recurring security problems. They are classified into various categories, each containing a set of attributes. However, the large number of patterns and classification schemes makes it difficult to choose a pattern for a given security problem. To apply patterns effectively, there must be a systematic method of organizing the patterns, so that it is possible to look up a design pattern unambiguously according to its purpose. While a lot of research focuses on developing new patterns and classifications, these issues have not been adequately addressed. In this paper, we present a novel approach of applying Formal Concept Analysis (FCA) on a chosen set of patterns classified according to a common set of attributes. The resulting concept lattice can be used for mining knowledge from the concepts, identifying pattern groups, and their relationships with the goal of applying appropriate patterns to security requirements. We propose the use of FCA over conventional data analysis methods for the simplicity of data preparation, the discovery of hidden knowledge, and cluster interpretation, with a visual representation of the pattern domain.

Keywords: Design patterns, Security, Concept lattice.

1. Introduction

Integrating security into all aspects of Information Systems has become critical. However, it requires skill and expertise to design secure systems. Expert knowledge in the form of design patterns can provide valuable guidance to the designers. The first design patterns for security were developed by Yoder and Barcalow in 1997 [1].

Subsequently, many patterns and pattern catalogues emerged. Due to a large
number of patterns, searching for a pattern becomes a daunting task, leading to the poor applicability of patterns in practice. This requires patterns be well organized. Researchers have proposed several classification schemes, each with multiple attributes. However, as more security threats and vulnerabilities were discovered, more patterns and newer classification methodologies were identified. Existing patterns need to be reclassified and retrofitted into the new schemes, which is a challenging task. Pattern applicability will improve if all patterns are classified uniformly using the same attributes. In the present work, we have identified a set of attributes which can be applied to all patterns.

Formal Concept Analysis (FCA) is an analytical method on tabular data which captures the relationships in a concept lattice [2]. We propose a novel approach of applying FCA to the classified pattern dataset. This will enable knowledge discovery and effective navigation through the security design pattern landscape. Patterns which can handle one or even multiple security issues at a time can be easily identified. The concept lattice also gives a nice visualization of the pattern clusters.

The paper is organized as follows. Section 2 gives a brief background on security design patterns and their classification. It also discusses the theoretical concepts behind FCA. The steps carried out in applying FCA are discussed in Section 3. The results are discussed in Section 4, followed by a conclusion in Section 5.

2. Background

In this section, we present a brief background on security design patterns and FCA.

2.1. Security Design Patterns

Alexander et al. first introduced the concept of design patterns for use of living spaces, in his book ‘A Pattern Language’, published in 1977 [3]. The concept of a pattern was adopted by the software community in the book ‘Design Patterns:
Elements of Reusable Object-Oriented Software’, published in 1994 whose authors Gamma, Helm, Johnson, and Vlissides are known as the ‘Gang of Four’ [4]. The first design patterns related to security were published by Yoder and Barcalow in 1997. Their paper described 7 design patterns. Subsequently, many design patterns and pattern catalogs emerged. Important pattern contributors are Romanosky [5], Kienzle et al. [6], Blakley and members of the Open Group security Forum [7], Hafiz et al. [8, 9] and the Microsoft Patterns and Practices group [10, 11]. In their book published in 2005, Schumacher and a working group of security pattern experts discussed 46 patterns [12]. Steel et al. [13] compiled a catalog of 23 core security patterns for J2EE and web services. In their guide ‘Security Design Patterns’, Dougherty et al., at the Carnegie Mellon Software Engineering Institute, described 15 design patterns [14]. Fernandez presented 68 design patterns in the book ‘Security Patterns in Practice: Designing Secure Architectures Using Software Patterns’ [15]. Today the total number of security design patterns is around 400 [16, 17]. To organize the patterns, several classification schemes have been developed [17, 18]. Each classification scheme contains a set of criteria or attributes. Some important classification schemes and their criteria are listed in Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Attributes/Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrogatives [19]</td>
<td>Purpose (Why), Data (What), Function (How), Timing (When), Network (Where), People (Who) and Scorecard (Test)</td>
</tr>
<tr>
<td>Purpose [16]</td>
<td>Structural, Behavioural, Generic, Procedural, Creational</td>
</tr>
<tr>
<td>Location [21]</td>
<td>Core, Perimeter, Exterior</td>
</tr>
<tr>
<td>Logical tiers [16]</td>
<td>Application, Host, Client, Logic, Data, Integration, Platform and OS, Distribution, Transport, Network</td>
</tr>
<tr>
<td>Threat Model [22]</td>
<td>Spoofing, Tampering, Information Disclosure, Repudiation, Escalation of Privilege, Denial of Service</td>
</tr>
<tr>
<td>Constraint [22]</td>
<td>Regulatorly, Organizational, Human, Mechanism</td>
</tr>
<tr>
<td>Response [22]</td>
<td>Avoidance, Deterrence, Prevention, Detection, Mitigation, Recovery, Forensics</td>
</tr>
<tr>
<td>Domain [16]</td>
<td>Web Services, J2EE, Microsoft</td>
</tr>
</tbody>
</table>
2.2. Theoretical basis of formal concept analysis

Formal Concept Analysis is also known as Galois lattice mining and is a branch of lattice theory that allows identification of meaningful groups of objects that have common attributes [24, 25]. According to Bělohlávek [26], the advantage of using FCA is that hidden concepts and dependencies in data can be discovered and reasoned with. The concepts and dependencies can also be visualized. FCA has been applied to data for knowledge discovery in various fields, such as biology, semantic web, machine learning, medical information systems, decision systems and others [27, 28].

2.2.1. Context

A context is a triple: $C = (E, P, I)$ where $E$ is a finite set of elements (objects), $P$ is a finite set of properties (attributes) and $I$ is a binary relation between $E$ and $P$ (such that $I \subseteq E \times P$). $I$ is called the formal context and represented as a table. An example is shown in Fig. 1.

![Fig. 1. Formal context.](image)

The rows in the formal context are the objects and columns are the attributes. Each table entry is Boolean, which indicates whether the object has that attribute. FCA is a data analysis method where data is the formal context. Each formal context is transformed into a mathematical structure called a concept lattice. According to Wille [24], concepts transform information into knowledge.

2.2.2. Concept

A concept is a pair of sets $(X, Y)$ where $X$ is a set of elements called the extent and $Y$ is a set of properties called the intent; given : $X \subseteq E$ and $Y \subseteq P$, $Y = \sigma(X)$ and $X = \pi(Y)$ where $\sigma(X) = \{p \in P \mid \forall e \in X : (p, e) \in I\}$ and $\pi(Y) = \{e \in E \mid \forall p \in Y : (p, e) \in I\}$ [29].

Therefore a concept is a maximal collection of elements sharing common properties, i.e., any $e \in E$ has all properties in $P$, and all properties $p \in P$ fit all elements in $E$. From the formal context shown in Fig. 1, $\{\{o_1, o_3, o_4\}, \{a_1\}\}$ is a concept. However, $\{\{o_1, o_3\}, \{a_1\}\}$ is not a concept since, although $\sigma(\{o_1, o_3\})=\{a_1\}$, $\pi(a_1) = \{o_1, o_3, o_4\}$. Similarly $\{\{o_1, o_3\}, \{a_1, a_3\}\}$ is a concept since $\sigma(\{o_1, o_3\})=\{a_1, a_3\}$ and $\pi(a_1, a_3) = \{o_1, o_3\}$.

FCA produces three kinds of output from the input data:
i. Concept lattice

The formal context is visualized using a concept lattice or Hasse diagram [29]. In a Hasse diagram, the nodes represent concepts. The lattice is a collection of formal concepts, which are hierarchically ordered by a subconcept-superconcept relationship. A concept \( (A_1, B_1) \) is a subconcept of \( (A_2, B_2) \) if \( (A_1, B_1) \subseteq (A_2, B_2) \) (iff \( B_2 \subseteq B_1 \)). This indicates that \( (A_1, B_1) \) is more specific than \( (A_2, B_2) \). The concept lattice is represented graphically to support analysis, mining, visualization and interpretation. The concept lattice for the formal context in Fig. 1 is shown in Fig. 2.

![Concept lattice diagram](image)

**Fig. 2. Concept lattice.**

ii. Attribute implications

An attribute implication describes a particular dependency which is valid in the data. It indicates the relationships between attributes. It is an expression of the form: \( A \Rightarrow B \) where \( A \) and \( B \) represent a set of attributes such that if an object has all attributes from \( A \), it has all attributes from \( B \) as well. For example, for the formal context shown in Fig. 1, \( a_3 \Rightarrow a_1 \) is an implication, since an object which has attribute \( a_3 \), also has attribute \( a_1 \). The attribute implications of the above formal context are:

- \( a_1 \, a_2 \Rightarrow a_3 \);
- \( a_3 \Rightarrow a_1 \);
- \( a_4 \Rightarrow a_2 \);

iii. Association rules

An association rule is an implication expression of the form \( X \rightarrow Y \) where \( X \) and \( Y \) are disjoint itemsets. The strength of an association rule is measured in terms of its Support and Confidence. Support determines how often the rule applies to the dataset. Support, \( \sigma(X \rightarrow Y) = \eta(X \cup Y) / N \) where \( N \) is the number of objects in the set. Confidence determines how frequently items in \( Y \) appear where \( X \) appears in a rule. Confidence, \( c(X \rightarrow Y) = \eta(X \cup Y) / \eta(X) \). Association rules that have a higher support are interesting because they reveal important characteristics of the
data set. Confidence is a measure of the reliability of the inference. For example, the rule a3=>a1 has a support of 0.5 and confidence of 100%.

When the formal context is very large, the concept lattice is also very large and it is unreasonable to display the complete data in a single lattice diagram. In such a case, the following approaches may be used:

a) Reduce the number of objects or attributes by observation.

b) Split the data, carry out a separate analysis, and merge the results.

c) Calculate the iceberg lattice of the context.

2.2.3. Iceberg concept lattice

The set of all frequent concepts of a context is called the iceberg lattice of the context. It consists of the top-most concepts of the lattice, which provide the most global structuring of the domain [2, 30]. A concept is called a frequent concept if its intent is frequent. These concepts are also called an order filter or upset. Let minsupp be a threshold ∈ [0,1], then B is said to be frequent itemset if σ(B) >= minsupp [30]. Figure 3 shows the iceberg lattice with the concepts having a minimum support (minsupp) = 0.5.

![Iceberg lattice](image)

Fig. 3. Iceberg lattice.

3. Using FCA For Pattern Data Mining

The steps that are followed for applying FCA on security pattern data are shown in Fig. 4.

![Steps in applying FCA](image)

Fig. 4. Steps in applying FCA.

It involves three stages:
a) Data preparation: The activities in this stage are building the pattern list, selecting the attributes for classification and constructing the formal context. The formal context is a matrix of pattern-attribute binary relations.

b) FCA: This stage includes the tasks of choosing the FCA tool, constructing the concept lattice, extracting the concepts, and generating the iceberg lattice based on minimum support.

c) Post-processing: The data that is generated by FCA is used for querying and mining useful information.

3.1. Data preparation

The pattern classification methodologies and their criteria are listed in Table 1. Multiple classification schemes make it difficult to apply them to all patterns. A core set of criteria can aid efficient organization of the pattern landscape. We propose that applying a classification scheme that is based on fundamental core security principles [31] along with the threats [32] that can be mitigated by the pattern, will help designers to apply patterns more effectively. Moreover, these criteria can be applied to most of the security patterns. Table 2 lists the 13 security criteria which are applied to the security patterns.

<table>
<thead>
<tr>
<th>No.</th>
<th>Attribute / Description</th>
<th>Security Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Authentication / Identity</td>
<td>Identifies the principals - users, machines, and processes.</td>
</tr>
<tr>
<td>2.</td>
<td>Access Control / Authorization</td>
<td>Determines which principal may access which data, i.e., it defines the access privileges.</td>
</tr>
<tr>
<td>3.</td>
<td>Integrity</td>
<td>Ensures that data and communications will not be compromised by active attacks.</td>
</tr>
<tr>
<td>4.</td>
<td>Confidentiality</td>
<td>Ensure data and communication privacy from unauthorized access.</td>
</tr>
<tr>
<td>5.</td>
<td>Availability</td>
<td>Assures that authorized users can use the resources when they are required.</td>
</tr>
<tr>
<td>6.</td>
<td>Non-Repudiation</td>
<td>Users’ refusal to acknowledge participation in a transaction.</td>
</tr>
<tr>
<td>7.</td>
<td>Accountability</td>
<td>Logging certain actions for audits.</td>
</tr>
<tr>
<td>8.</td>
<td>Key management</td>
<td>Related to the generation, sharing and storage of keys.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Threats</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9.</td>
<td>Spoofing</td>
</tr>
<tr>
<td>10.</td>
<td>Tampering</td>
</tr>
<tr>
<td>11.</td>
<td>Information Disclosure</td>
</tr>
<tr>
<td>12.</td>
<td>Denial of Service</td>
</tr>
<tr>
<td>13.</td>
<td>Elevation of privilege</td>
</tr>
</tbody>
</table>
The 13 attributes, a set of 192 security design patterns, and the binary relation between them forms the formal context for FCA. The attributes are represented as columns and patterns as rows. Each table element is assigned a binary value - 1 or 0 depending on whether the pattern satisfies that criterion or not. A representative sample is shown in Table 3.

<table>
<thead>
<tr>
<th>No.</th>
<th>Pattern Name</th>
<th>Authentication</th>
<th>Access Control</th>
<th>Integrity</th>
<th>Confidentiality</th>
<th>Non Repudiation</th>
<th>Availability</th>
<th>Accountability</th>
<th>Key Management</th>
<th>Spoofing</th>
<th>Tampering</th>
<th>Information Disclosure</th>
<th>Denial of Service</th>
<th>Elevation of Privilege</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Pattern for WS-Trust</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Access Control List</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192</td>
<td>XML Encryption</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

3.2. Applying FCA

The pattern-attribute binary relationship table is the input to the FCA tool. There are a variety of software tools available for FCA [33]. The formal context is represented in a format as specified by the FCA tool (.csv is commonly supported). The tools used to carry out the present research work are ConExp and Galicia. The following outputs are obtained as results:

i. Concepts and concept lattice.

ii. Implications and Association Rules.

iii. Iceberg lattice.

4. Results and Discussion

The concept lattice that is generated by FCA for the pattern dataset is shown in Fig. 5. The height of the lattice is 8 and there are 148 concepts in the lattice. Each node in the lattice represents a concept \{I,E\}, where I represents the intent i.e. the attributes and E represents the extent, i.e., the patterns in the concept. The patterns in a single group share common attributes. Pattern clusters at a higher level in the lattice support fewer attributes than those at lower levels. These patterns are more generic and solve fewer security issues. For example, the pattern group \{Trust Partitioning, chroot jail\} has only one attribute: \{Tampering\}. Concepts at lower lattice levels have more attributes and hence, the patterns in these concepts are more useful. These patterns address multiple issues. For example, the lower level group \{Multilevel Security, Multilevel Secure Partitions\} has the attributes \{Confidentiality, Information Disclosure, Access Control, Tampering, Integrity, etc\}.
Elevation of Privilege]. The lattice gives a schematic of the entire security pattern domain and clearly brings out the relationships between the patterns and attributes.

4.1. Concepts

Each node of the lattice represents a concept which is connected to other concepts by edges. The node in Fig. 6 corresponds to the concept having the intent \{Integrity, Tampering\} and an extent of 53 patterns. This concept has 6 sub-concepts, having intents as: \{Integrity, Tampering, Access Control\}, \{Integrity, Tampering, Availability\}, \{Integrity, Tampering, Escalation of Privilege\}, \{Integrity, Tampering, Accountability\}, \{Integrity, Tampering, Authentication\} and \{Integrity, Tampering, Information Disclosure\}. It has two superconcepts which have the attributes of \{Integrity\} and \{Tampering\} respectively.
The concepts which have maximum number of attributes in their intent are shown in Table 4. Out of the 148 concepts in the lattice, concepts 1 and 2 have 7 attributes each, while the others have 6 attributes each. The patterns in these concepts are more useful patterns because they are capable of handling more number of security issues. The superconcept-subconcept relationship is clearly visible in the concepts. For example, 1 and 2 are subconcepts of 3.

Table 4. Concepts with maximum attributes.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Information Disclosure, Confidentiality, Integrity, Access Control, Tampering, Authentication, Spoofing (Total 7)</td>
<td>Capability, Check Point, Secure Visitor, Secure Process Thread, Limited Access With Errors</td>
</tr>
<tr>
<td>2. Access Control, Tampering, Integrity, Authentication, Confidentiality, Spoofing, Accountability (Total 7)</td>
<td>Full Access With Errors</td>
</tr>
<tr>
<td>5. Information Disclosure, Confidentiality, Integrity, Authentication, Access Control, Escalation of Privilege (Total 6)</td>
<td>Access Session</td>
</tr>
<tr>
<td>6. Tampering, Integrity, Authentication, Confidentiality, Spoofing, Accountability (Total 6)</td>
<td>Full Access With Errors, Password Authentication</td>
</tr>
<tr>
<td>7. Spoofing, Authentication, Information Disclosure, Confidentiality, Accountability, Non-Repudiation (Total 6)</td>
<td>A pattern for WS Trust</td>
</tr>
<tr>
<td>8. Tampering, Information Disclosure, Confidentiality, Integrity, Accountability, Non-Repudiation (Total 6)</td>
<td>Secure Logger</td>
</tr>
</tbody>
</table>
4.2. Iceberg lattice

All the concepts in the lattice cannot be easily viewed at a time. It is difficult to draw any inferences just by observation. Hence, additional tools and algorithms are used to extract lattice information. An iceberg lattice is the set of all frequent concepts of the context, whose support meets a given threshold. It encapsulates the common knowledge, condensed representation or most generic values of the context. The iceberg lattice for minsupp=0.4 is shown in Fig. 7.

As we can see, this lattice has only 3 concepts. This indicates that these three concepts form the topmost concepts of the lattice. These concepts are:

- Confidentiality which has an extent of 81 (42%)
- Integrity which has an extent of 91 (47%)
- Information Disclosure which has an extent of 80 (42%)

The inference from the iceberg lattice is that out of the 192 patterns in the set, more than 40% of the patterns satisfy the criteria of Confidentiality, Integrity, and Information Disclosure. There is no other attribute having a support of more than 40%. Further reduction of minsupp to 0.2 yields the iceberg lattice shown in Fig. 8. It has 11 concepts and the additional concepts are: Access Control (extent: 53, support 28%), Tampering (extent: 60, support 31%) Authentication, Spoofing, Authentication+ Spoofing (extent: 39, support 20%), Confidentiality+ Information Disclosure (extent: 55 objects, support 29%), Confidentiality+Integrity (extent: 42 objects, support 22%), Integrity+Tampering (extent: 53 objects, support 28%).

4.3. Implications

The concepts and their hierarchical structure reveal the relationships between patterns and attributes. The implications of the context give valuable insights into attribute relationships. This can be mined to extract useful information regarding attributes. An attribute implication describes an attribute dependency which holds in the data; for example, concepts which have the attributes of Access control, Tampering, and Information Disclosure also have the attribute of Integrity. A part of the lattice is shown in Fig. 9.

From the diagram, we can detect the implication Access Control, Spoofing => Authentication. It is indicated by the fact that there is no concept having ‘Access Control’ and ‘Spoofing’ in its intent, but not ‘Authentication’. 23 patterns in the data set have this implication and hence, a support of 12%.
5. Conclusions

Formal Concept Analysis is a useful tool to visualize the concepts within data and the relationships between concepts. It is a data analysis method which enables the discovery of hidden knowledge existing in the data. In this paper, we have applied FCA to a set of 192 security design patterns which have been classified according to 13 criteria. The criteria are the fundamental security principles and security threats. The resulting concept lattice gives valuable information about pattern clusters and the attributes of the cluster. Practitioners can easily identify patterns which solve specific security issues. The grouping of patterns indicates the relationships between patterns and can be used to identify redundant patterns. Since the number of concepts is large, it is not possible to draw inferences from the lattice just by observation. Additional inferences are drawn from the iceberg lattice, which gives...
the topmost concepts in the lattice hierarchy. These concepts form the most generic concepts of the lattice. Implications and association rules reveal relationships and dependencies between the attributes. Support and confidence measures indicate the strength and reliability of the association rules. Future work includes association rule mining and a detailed exploration of the concepts. The concepts and cluster knowledge, if made readily available in the form of a web-based tool, will enable pattern search based on security goals and thereby improve pattern applicability.

References


