Data-Driven Chinese Walls

Gulsum Akkuzu
School of Computing, University of Portsmouth, United Kingdom
gulsum.akkuzu@port.ac.uk

Benjamin Aziz
School of Computing, University of Portsmouth, United Kingdom
benjamin.aziz@port.ac.uk

Abstract

Security policy and access control models are often based on qualitative attributes, e.g. security labels, cryptographic credentials. In this paper, we enrich one such model, namely the Chinese Walls model, with quantitative attributes derived from data. Therefore, we advocate a data-driven approach that considers a quantitative definition of access we term, working relations.

2012 ACM Subject Classification Security and privacy → Access control

Keywords and phrases Access Control, Big Data, Security Policies, Chinese Walls Model

Digital Object Identifier 10.4230/OASIcs.ICCSW.2018.3

Category Main Track

1 Introduction

Organisations require controlling and monitoring the access of their information, shared files and resources in order to ensure the security of information and assets. Access control provides access rights to authorized users. Unauthorized users are refused to access to information and assets. There are various access control models in the literature, including mandatory access control, discretionary access control, role-based access control, lattice-based information flow control and Chinese Wall Access Control (CWAC) [4, 10].

CWAC was introduced by Brewer and Nash [2] with the aim preventing information flows that cause Conflicts of Interest (CoI) for consultants. It is considered in the commercial domain in which consultants and analysts of organisations access sets of data resources from different groups in companies that provide different types of services. CWAC prevents data leaks from one company to another within the same CoI class.

CWAC, like many other models, lacks quantitative attributes and analysis that could render the model more usable within data-driven domains. We therefore argue that such a model should be enhanced with definitions that take into consideration statistical information derived from available datasets. We introduce in this paper one such initial enhancement by taking into account working relations of users towards computers they access. A working relation represents persistent accesses that exceed some minimum number of times. We define a new version of the simple secrecy property based on working relations. To the best of our knowledge, this research is the first for incorporating quantitative aspects such as probabilistic or stochastic variations of CWAC model.

2 Related Work

Our work is closely related to the basic concept of CWAC. We provide an overview of the most relevant works to the this paper in the followings. CWAC is a well known access control for the secured sharing of commercial consultancy services. A concept analysis is introduced which shows effects of lattice structure to classify the access permission of consultants based on the CWAC. And also helps to understand the CWAC access permission of every consultant depends on their level in the lattice structure [11].

A new way of thinking on Chinese Wall Model Security has been brought by Cheng and et all [3], they address risk access control decisions on the CWAC and other security policies where access decisions are based on the history. The new concept has been illustrated as a Fuzzy Logic control system because of the uncertainty in security labels, risk modulating factors and loss variance based access decision have effective roles on the access decisions.

CWAC is proposed to address the CoI problems in the decentralized work-flow environment [13] . The modified CWAC for decentralized control that solves the various problems such as, placing the task execution agents in same COI class's, undesirable results that are similar and tend to be in the wrong side of the Chinese Wall.

Formal Concept Analysis CWAC is modelled [12] , results show that the proposed method is able to satisfy the limitation of the Chinese Wall Security Policy and its main properties that are simple security (ss) and *-property. Similarities between CWAC security properties and Multi-Level Security (MLS) policies' security attributes are described by McLean [8]. Bella-LaPadula is one of the MLS policies, it has a simple security rules and *-property that are partly similar to CWAC access rules .According to Gollmann [5], ss-property allows a subject s to access an object o either only if object o has already been accessed by the user or request is entirely different conflict of class. He describes *-property for restricting write access; an object o can be accessed by a subject s only if s/he has no read access to any conflict class object o'. All of the previous works mentioned so far, has not considered a quantitative approach on Chinese Wall Security Model. We believe that our work is the only work that redefines security formula of CWAC and approaches quantitative aspects of it.

3 An Overview of Chinese Wall Security Model

CWAC is one of the most common policy that is usually used in commercial organizations [2]. The main idea behind this is that one is not allowed to access to two conflicting classes or competitor organizations' files at the same time. It can be described quite simply, object that is the wrong side of the wall should not been allowed to be accessed by a subject.

CWAC has two access restriction properties; simple security rule and *-property. Simple security rule gives an access a subject to an object if the object either is in the same company datasets and object already accessed by subject or belongs to a completely different conflict of interest class. Chinese Wall *-property comprises permissions on write access, it is permitted if access is permitted by the simple security rule and object can not be read which is in different organisation dataset than the one which write access is requested for [5].

Our model has similarities with the exist CWAC on the checking of the historical accesses but also we have a new approach that is working relationship. In our case, if a user has accessed to a computer more than n times then we think user has a working relationship with that computer, hence if the user attempts to access a computer that belongs to the CoI, then our property does not allow user. We explain our new model definitions in the following section and we implement our new model on the real world dataset in Section 5.2.

4 A New Data-driven Chinese Walls Model

The existing Chinese Walls model [2] is based on the concept that a subject may have accessed some object in the past that belonged to a conflicting organisation. However, the model is coarsely defined as it relies on the single notion of access. There are no attempts in literature to provide more refined definitions that would incorporate quantitative aspects such as probabilistic or stochastic variations of this model. Therefore, we start with the definition of a *working relation* between a subject and an object to capture such quantification in access control. We start first by reviewing the basic concepts underlying the Chinese Walls model.

Subjects are defined as the set $s, s' \ldots \in \mathcal{S}$. These could be users or computer machines. Objects are defined as the set $o, o' \ldots \in \mathcal{O}$. These could be individual files. Companies (represented by their datasets) are defined as the set $c, c' \ldots \in \mathcal{C}$. There are two labeling operations on objects. The first is $y : \mathcal{O} \to \mathcal{C}$, which defines for an object the company (dataset) to which it belongs. The second labeling operation is $x : \mathcal{O} \to \wp(\mathcal{C})$, which defines for an object, the set of companies (datasets) that are in conflict with its owner. We call this set, the *conflict class* for o.

Based on the above concepts, the history of subjects' access to objects is defined as a matrix $N : \mathcal{S} \times \mathcal{O} \to \mathbb{B}$, where $N_{s,o} = \textbf{True}$ means that s has accessed o in the past, and $N_{s,o} = \textbf{False}$ means it has not. However, in our model, we redefine the history of accesses as a numeric concept rather than a Boolean one.

▶ **Definition 1** (Numeric History of Accesses)**.** We define the numeric history of accesses of objects by subjects as a matrix, $N : \mathcal{S} \times \mathcal{O} \to \mathbb{N}$, which returns for each subject a natural number, n, representing the number of times a subject s has accessed an object o in the past:

$$N_{s,o} = n$$

Using this new definition of the history of accesses, we can define a new relation to capture working relationships between users and computers.

▶ **Definition 2** (Working Relations)**.** We say that a subject s (e.g. a user) has a *working relation* with an object o (e.g. a computer) written as a predicate, $wr(s, o)$, if and only if, based on some predefined minimum number of accesses, n, then:

$$N_{s,o} \geq n$$

In other words, s, in its history, has accessed o at least n number of times. The choice of n depends on the organisation or on the context in which the policy is deployed. By contrast, the standard case where s is deemed to have only *accessed* o, is represented by the next definition.

▶ **Definition 3** (Standard Access)**.** We say that a subject s (e.g. a user) has *accessed* an object o (e.g. a computer) if and only if, based on some predefined minimum number of accesses, n, then:

$$n > N_{s,o} \geq 1$$

In other words, s, in its history, has at least accessed o once, but fewer times than n. Therefore, s has accessed o but not have had a working relation with o. Finally, the case of no access is defined simply as as follows.

```
1, U1, C1
1, U1, C2
2, U2, C3
3, U3, C4
6, U4, C5
7, U4, C5
12, U8, C9
```

Figure 1 Example data lines.

▶ **Definition 4** (No Access)**.** We say that a subject s (e.g. a user) has *not accessed* an object o (e.g. a computer) in its history, if and only if, the following holds true:

$$N_{s,o} = 0$$

We now introduce the new variation of the Chinese Wall accessibility property, based on the definition of a working relation above.

▶ **Property 1** (Working Relations-based Simple Security)**.** *A subject, s, can access an object, o, if and only if, for all objects o', it must be the case that:*

$$(wr(s, o') = \textbf{\textit{True}}) \;\Rightarrow\; (y(o) = y(o') \;\vee\; y(o) \notin x(o'))$$

This property weakens the original Simple Security property defined by Brewer and Nash [2] in that it takes into account only that part of the history of the subject where accesses to objects reached to some particular level of significance (i.e. n as defined in Definition 1). Informally, this means that we are not worried with subjects who have accessed objects fewer times that n in their history.

5 Case Example: LANL Dataset

5.1 Dataset Description

We use here the "User-Computer Authentication Associations in Time" (UCAAT) dataset [6, 9] collected by the Los Alamos National Laboratory (LANL) [1], as our case example to validate the ideas presented in the previous section. We used first five thousands data from the dataset due to memory of the computer that is used for coding. The data ranges over 9 months and represents 708,304,516 successful authentication events from users to computers. An example of some lines in the dataset is shown in Figure 1.

Each line contains three metadata elements; the first represents the time at which the authentication event occurred, the second represents the user who logged in into the computer, and the third the computer on which the login happened. The time epoch starts at 1 with a resolution of 1 second. There are in total 11,362 users, represented by the pseudo values `Ui` and 22,284 computers, represented by the pseudo values `Cj`, where `i, j` represent the number of the user and the computer, respectively. To enhance anonymity, the time frame of the actual data collection is not provided and some centralized computers (e.g. active directory servers) and their associated authentication events have also been removed. The dataset is available either as a single compressed file (size 2.3GB) or as a set of 9 individual files (sizes ranging from 177MB to 273MB).

5.2 Model Implementation Based on the Dataset

We now instantiate our new data-driven Chinese Walls model using data from the UCAAT dataset. Due to the size of the dataset, we selected as a proof-of-concept only the first 5000 entries. Initially, we find the number of times each user logged in to a specific machine. Table 1 shows an example of such analysis.

Table 1 An example table showing the number of times users log in to computers.

User Name	Computer Name	Number of Login Times
U1	C1	2
	C2	4
	C978	2
U12	C54	62
	C94	8
	C801	3
U66	C1	18
	C117	3
	C133	1
U105	C113	2
	C130	3
	C160	37
U106	C136	8
U116	C155	32
U127	C155	41
U13	C14	21
	C172	20
	C282	20
	C32	30

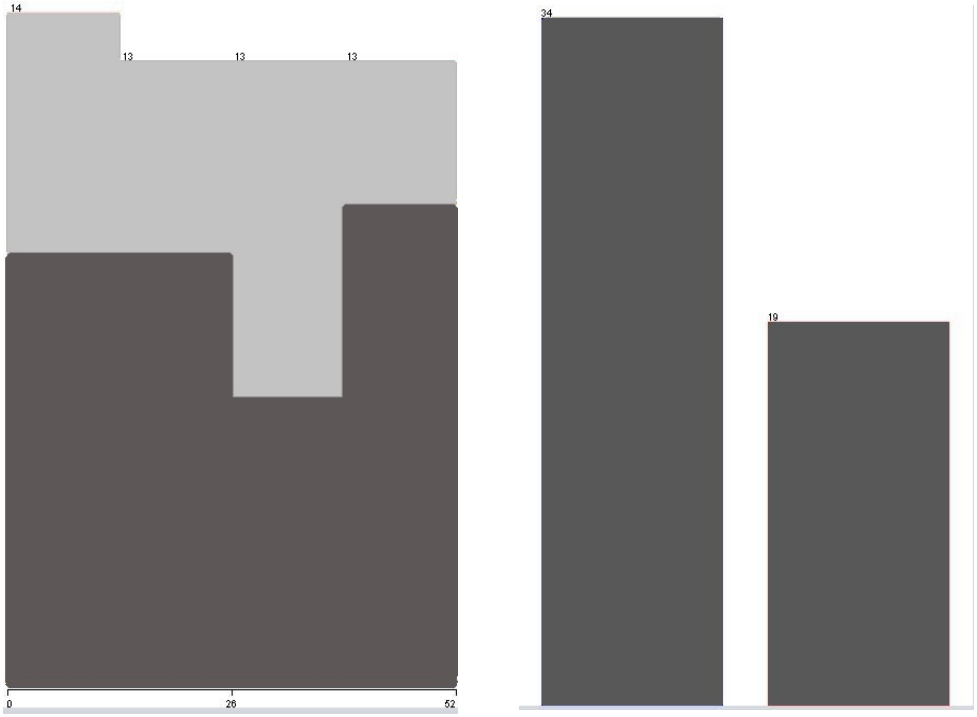
We applied clustering techniques to find the value of n (for definition of n see Section 4. Clustering is a technique that is used to group data together [7]. It is used to classify each data point into a specific group. We give the number of login time, and classify them into two groups. One of the common clustering algorithms is K-means which is a method to partition a data set into k groups [14]. The clustering results are given in Figure 2. The numbers are clustered into two groups and the boundary number of the numbers was 20. Therefore, in our case (for first five thousands data) n value is 20. We use u instead of s, similarly c instead of o. In this way, if

$$N_{u,c} \geq 20$$

then we say user has a working relations with the computer. For example, U12 has working relations with C54 because U12 has logged into C54 68 times. Table 2 represents the full list of users with working relations to computers. Standard access value becomes as follows;

$$20 > N_{u,c} \geq 1$$

For example, U1 has only standard access in relation to all the computers that have been logged in to by U1, because $N_{u,c}$ is between 1 and 20. On the other hand, U13 has working relation with all computers (see Table 1), n is greater than 20. We use the same approach with the Chinese Walls model *simple security*; In our case, if a user wants to login to the computer and s/he has a working relation with a CoI group of computers, then our model



(a) Clustering 1.

(b) Clustering 2.

Figure 2 K means Clustering.

Table 2 The group of users who have working relations with computers.

User Name	Computer Name	User Name	Computer Name
U12	C54,C13	U105	C160
U116	C155	U120	C164
U124	C168, C192	U127	C155
U128	C167, C176	U13	C14, C172, C282, C32, C42
U130	C179	U153	C207
U156	C161	U159	C160
U16	C17, C148	U179	C175
U184	C154	U188	C256
U197	C269	U202	C275
U204	C101	U21	C22
U211	C286	U216	C291
U29	C30	U39	C41
U53	C58	U6	C7
U60	C108, C173	U63	C234, C68
U66	C1	U67	C49
U68	C71	U77	C234, C78
U93	C107,C53	U92	C154
U95	C154	U97	C161
U105	C160	U116	C155
U120	C164	U9	C10
U96	C115	U97	C116, C161

does not allow her/him to access the computer. A user, s, can log into a computer, o, if and only if, for all computers o', it is the case that:

$$(wr(u, c') = \textbf{True}) \Rightarrow (y(c) = y(c') \vee y(c) \notin x(c'))$$

The property allows a user if and only if, s/he has either standard access or, has no access which means that s/he has not logged-in to the computers. However, if a user has working relation, then s/he can not be permitted access to computers that belong to a conflict class. Table 2 represents the full list of users with working relations to computers.

6 Conclusion and Future Work

In this paper, we first defined a robust, quantitative model for Chinese Walls Access Control. We showed that our new model is suitable to apply on the real world dataset, in our case, we used the UCAAT dataset for implementing our quantitative Chinese Walls Model. We then explained the way to find out the value of the n that shows boundary of access times by applying clustering algorithms. Our model allows to put access restrictions on the sensitive and personal information so that users cannot manipulate other users' sensitive files for their own advantages. There is an opportunity for the further work, we plan to extend this work with quantitative description of *property. We also believe that our model can be applied other Multi Level Security (MLS) models.

References

1 Los Alamos National Laboratory: Cyber Security Science. `https://csr.lanl.gov/data/`. Accessed: 14-06-2018.

2 D.F.C. Brewer and M.J. Nash. The Chinese Wall Security Policy. In *Proceedings of the* 1989 *IEEE Symposium on Security and Privacy*, pages 206–214, Oakland, California, USA, 1989. IEEE Computer Society Press.

3 Pau–Chen Cheng, Pankaj Rohatgi, Claudia Keser, Paul A. Karger, Grant M. Wagner, and Angela Schuett Reninger. Fuzzy Multi-Level Security:An experiment on quantified risk-adaptive access control. In *2007 IEEE Symposium on Security and Privacy (S&P 2007), 20-23 May 2007, Oakland, California, USA*, pages 222–230. IEEE Computer Society, 2007. `doi:10.1109/SP.2007.21`.

4 Dhillon G. and G. Torkzadeh. Value-focused assessment of information system security in organizations. *Information Systems Journal*, 16(3), 293-314, 2015.

5 Dieter Gollmann. *Computer Security*. John Wiley & Son Ltd, 1999.

6 Aric Hagberg, Alex Kent, Nathan Lemons, and Joshua Neil. Credential hopping in authentication graphs. In *2014 International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*. IEEE Computer Society, 2014.

7 Anil K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

8 McLean John. The algebra of security. In *In Security and Privacy*, volume 1290, pages 2–7. IEEE Symposium, 1988.

9 Alexander D. Kent. User-Computer Authentication Associations in Time. Los Alamos National Laboratory, 2014. `doi:10.11578/1160076`.

10 Butler Lampson. A note on the confinement problem. *Communications of the ACM*, 16(10):613–615, 1973.

11 S. C. Mouliswaran, C. A. Kumar, and C Chandrasekar. Modeling Chinese wall access control using formal concept analysis. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*. IEEE, 2014. `doi:10.1109/IC3I.2014.7019619`.

12 S. C. Mouliswaran, C. A. Kumar, and C Chandrasekar. Modeling Chinese wall access control using formal concept analysis. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*. IEEE, 2014. doi:10.1109/IC3I.2014.7019619.

13 Atluri Vijayalakshmi, Chun Soon, and Mazzoleni Pietro. A Chinese wall security model for decentralized workflow systems. In *CCS 2001, Proceedings of the 8th ACM Conference on Computer and Communications Security, Philadelphia, Pennsylvania, USA, November 6-8, 2001.*, pages 48–57. ACM, 2001. doi:10.1145/501983.501991.

14 Kiri Wagstaff, Claire Cardie, Seth Rogers, and Stefan Schrödl. Constrained k-means clustering with background knowledge. In Carla E. Brodley and Andrea Pohoreckyj Danyluk, editors, *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001*, pages 577–584. Morgan Kaufmann, 2001.