Biba Security Model Inspired Social Media Security Controls

Mike Westmacott
Thales UK
Reading RG2 6GF
mike.westmacott@uk.thalesgroup.com

Abstract

This paper presents a theoretical mechanism for managing anti-social and abusive behaviours in a subset of social media applications. The mechanism adapts the Biba security model, itself a variant of the Bell-Lapadula security model, replacing integrity with a security characteristic of 'identifiability'. The new model states that a subject with a high level of identifiability may not read data (content) from a subject with a lower level of identifiability, and that subjects with low levels of identifiability may not write (send content) to those with a higher level. It also states that discretionary access may be granted. The desirability, feasibility and viability of this mechanism are discussed.

1 Background

A core issue with harassment, grooming, libel, and other anti-social online behaviours in both closed and open online environments is that of anonymity. This leads directly to the lack of a means to attribute such behaviour to individuals and their actions. Hlavach and Freivogel (2011) reported that there is a relationship between anonymity and increased abusive behaviour in online news comments, and the concept of deindividuation was explored in the 1960s by Zimbardo (1969) of Stanford University and others.

This proposal attempts to deal with the issues of harassment, grooming, libel, and other anti-social online behaviours. The premise is that a platform that provides strong identity (user's accounts tied strongly to information that permits attribution of activity to a legal person) can implement a security model that prevents (or limits) the adverse behaviours. Behaviour is restricted through the risk of litigation or prosecution associated with anti-social or criminal behaviour when messaging other users.

The Biba model (Biba, 1975), a variation of the Bell-Lapadula (Bell and LaPadula, 1973) model, is designed to protect integrity of data, and is composed of three properties surrounding the security characteristic of integrity:

- **The Simple Integrity Property** states that a subject at a given level of integrity must not read data at a lower integrity level (read up).
- **The * (star) Integrity Property** states that a subject at a given level of integrity must not write to data at a higher level of integrity (write down).
- **The Invocation Property** states that a process from below cannot request higher access; only with subjects at an equal or lower level.

Such an implementation is used where it is the integrity of information - such as in military command, where a high ranking officer would not read instruction written from a lower ranking of individual, however that high ranking officer may write instruction down to lower levels.

2 Proposed Model

This work employs the security characteristic of identifiability - that is the degree to which a user may be associated with a human individual. The intentions of the model are to ensure that users doe not receive abusive information (in the form of messaging) from individuals who have a lower degree of identifiability, and therefore cannot easily be held accountable for their actions. It also provides for the mechanism to allow the former category to be able to send information and messages to widespread audience - so long as they are of a lower degree of identifiability.

The new model describes the following three security properties:

- **The Simple Identification Property** states that a subject at a given level of identifiability may not read data at a lower level of identifiability.
- **The * (star) Identification Property** states that a subject at a given level of identifiability may not write to data at a higher level of identifiability.
- **The Discretionary Security Property** uses an access matrix to specify discretionary access control.

The characteristic 'identifiability' may be stated as ranging from entirely anonymous (a difficult problem to solve in any system) to the absolutely identifiable, using suitable sets of identifiers (Marx, 1999). The characteristic must be applicable only to the system in question, and must not be influenced nor modified by the use of external systems.

The first property would prevent a user from reading posts from users who were less identifiable, regardless of the intended recipient. The second property prevents a user of low identifiability from posting to users who are more identifiable. The third property states that it is possible to permit exceptions to these rules on a discretionary basis.

3 Desirabiliy, Feasibility, and Viability

This model is likely only desirable in SNS (Social Networking Sites) contexts where anonymity and minimal identifiability are resulting in antisocial and abusive behaviour. This would include at-risk communities such as networks for children, closed-group communities like online gaming, and high-profile social-media influencers. Two examples that serve to illustrate the model when implemented in a platform are Twitter and Instagram. Both platforms suffer from large volumes of abuse, ranging from fake accounts to child grooming. Popular celebrities, for whom identity is already confirmed (albeit outside the platform) may use such a system as it would permit them to reach out to a very wide audience of all users, but receive only messages from individuals who were willing to be held responsible for responses. Young Instagram users could be protected by enforcing controls until they were of an adult age, where they could select to opt out of such a scheme.

The technical implementation of this model is straightforward, requiring only the following: A technical user identity (username), a set of identification characteristics that are ranked, and a mechanism for comparing the level of identification from each user. This would allow the first two properties of the model to operate, whilst the discretionary matrix would require a system for permitting exceptions. For interoperability a common system of conveying identification metadata would be necessary, and mandatory in the case of the use of federated identity services.

The most significant challenge is the adoption of such a security scheme. A sponsor would be necessary to validate the system, and a set of test participants to assess the mechanics and overall impact. Another issue is that of adoption by the platforms themselves – the perceived loss of anonymity for users (even if it were gradual) might be seen to be significantly detrimental to the core business mission of the platforms.

4 Future Work

Since the technical feasibility of this model is viable, both the desirability and viability must be addressed. The author proposes that studies to measure the desire of users to support such a system are conducted, through questionnaire and interview, and that viability be determined by engaging with the provider of a small scale SNS where this solution may be of most benefit - such as a platform for children.

References

David E. Bell and Leonard J. LaPadula. 1973. Secure computer systems:Mathematical foundations. Technical Report 2547, MITRE, Bedford, MA, US.

Kenneth J. Biba. 1975. Integrity considerations for security computer systems. Technical Report 3153, MITRE, Bedford, MA, US.

Laura Hlavach and William H Freivogel. 2011. Ethical implications of anonymous comments posted to online news stories. *Journal of Mass Media Ethics*, 26(1):21–37.

Gary T Marx. 1999. What's in a name? Some reflections on the sociology of anonymity. *The Information Society*, 15(2):99–112.

Philip. G. Zimbardo. 1969. *The human choice: Individuation, reason,and order versus deindividuation, impulse, and chaos.* Stanford University, Department of Psychology.