Database Security, II

DATABASE SECURITY, II
Status and Prospects

Results of the IFIP WG 11.3 Workshop on
Database Security
Kingston, Ontario, Canada, 5–7 October, 1988

Edited by

CARL E. LANDWEHR
Chairman of IFIP WG 11.3
Naval Research Laboratory
Washington, D.C., U.S.A.

1989

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD · TOKYO

ELSEVIER SCIENCE PUBLISHERS B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Distributors for the United States and Canada:
ELSEVIER SCIENCE PUBLISHING COMPANY INC.
655 Avenue of the Americas
New York, N.Y. 10010, U.S.A.

Library of Congress Cataloging-in-Publication Data

```
IFIP WG 11.3 Workshop on Database Security (1988 : Kingston, Ont.)
   Database security II : status and prospects : results of the IFIP
  WG 11.3 Workshop on Database Security, Kingston, Ontario, Canada,
  5-7 October 1988 / edited by Carl E. Landwehr.
       p.   cm.
    Bibliography: p.
    Includes index.
    ISBN 0-444-87483-6 (U.S.)
    1. Data bases--Security measures--Congresses.  2. Computers-
  -Access control--Congresses.   I. Landwehr, Carl E.   II. IFIP WG
  11.3.  III. Title.
  QA76.9.D32I35  1988
  005.8--dc20                                              89-11737
                                                               CIP
```

ISBN: 0 444 87483 6

No responsibility is assumed by the publisher or by IFIP for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

Printed in the Netherlands.

PREFACE

With this second volume in a series reporting research and development results in the area of database security, IFIP Working Group 11.3 renews its commitment to help society manage and protect the information it entrusts to database systems. This volume collects the papers presented and discussed at the second meeting of WG 11.3, held at Kingston, Ontario, October 5-7, 1988. It is offered both to document progress and to stimulate researchers to pursue the challenges this research area offers.

Working Group 11.3 continues to use its list of open research questions to frame its efforts. At the final session of the meeting, progress was reviewed against the question list. This discussion, as well as summaries of the paper presentations and discussions that occurred throughout the meeting, are recorded in the final chapter of this volume. Readers seeking an introduction to the book as a whole thus might do well to start at the back.

Topics that generated significant interest at the meeting included the development of a proper framework for stating database system security requirements, and particularly how the notion of *roles* could best be applied. The application of object-oriented database systems to enforce security requirements stated in terms of role authorizations appears a promising avenue for further work. Characterization and control of the problems of aggregation and inference were also the subject of considerable effort and some progress in the past year.

The utility of establishing a common problem framework for testing proposed solutions to database security problems was also recognized at the meeting. The group tentatively agreed to use a hypothetical database of medical information for this purpose, since this application can involve complex yet intuitive requirements for secrecy, integrity, and availability. Such a database could require controlling access authorizations for a variety of roles including those of patients, doctors, nurses, pharmacists, epidemiological researchers, and insurers.

I thank the members of IFIP Working Group 11.3 for their contributions to the meeting and to this book. Glenn MacEwen of Queen's University and David Bonyun of AIT Corp. organized the meeting, and as Program Chair, Tom Haigh of Honeywell organized the refereeing and presentations. I thank them and all of the authors of the work reported here for their dedication, hard work, and good fellowship. They yielded a stimulating meeting and a volume of papers that will reward readers. Dan Thomsen and Tom Haigh took the primary responsibility for summarizing the discussions and deserve a special note of thanks for that effort. Queen's University sponsored the workshop. The Naval Research Laboratory continues to be generous in its support of my IFIP participation. Finally, I thank my family for their encouragement and support.

Carl E. Landwehr
Chairman, IFIP WG 11.3
Naval Research Laboratory,
Washington, D.C.

CONTENTS

Policy and Models

Object-Oriented Data Management Systems

Distributed DBMSs and Tools

MLS Database Design and Implementation

Inference and Aggregation

Summary of Proceedings

DATABASE SECURITY, II: Status and Prospects
C.E. Landwehr (Editor)
Elsevier Science Publishers B.V. (North-Holland)

SECURITY MODELS AND ENTERPRISE MODELS

John E. Dobson

Computing Laboratory
University of Newcastle upon Tyne
NEWCASTLE upon TYNE NE1 7RU
United Kingdom

John A. McDermid

Department of Computer Science
University of York
YORK YO1 5DD
United Kingdom

Arguments are presented that suggest that information flow models of security are inappropriate for representing the real security concerns of security problem owners. These needs are for the representation of the "flow of meaning" rather than the flow of information. Meaning cannot be modelled without an understanding of the enterprise of which the secure system is but a part. A method is presented of modelling an enterprise so that the security-relevant features of actions, information, and the roles of agents can be described and analysed, and which permits organisational security policies to be stated. Some applications of the enterprise modelling technique to security issues are briefly outlined.

1. INTRODUCTION

1.1. Background and Objectives

There have been a number of important and influential milestones in the development of secure systems processing classified data. We will loosely refer to this application domain as military security. One of the most important milestones has been the realisation that it is possible to produce security models which are applicable to a wide class of secure systems. The seminal work in this area is that of Bell and LaPadula [5].

Bell and LaPadula showed that it was possible to give a fairly simple state-based formalisation of security in terms of the passive objects (e.g. files and directories) held in a computer, and the subjects (e.g. programs and processes) which act upon them. Having produced an abstract model of the state of a computer system, Bell and LaPadula identified classes of operation which changed the system state, then specified the security policy in terms of changes to that state. This gave rise to the well-known **simple security** property that a subject cannot read an object above its clearance, and the **star** (*) property that a subject may not write to an object below its clearance. Bell and LaPadula gave both an abstract definition of the model and showed how to interpret the model for real computer systems, e.g. Multics [5].

Most military secure systems developed since the late 1970's have been designed and built to the spirit of the Bell and LaPadula model, if not to the letter. The Bell and LaPadula model has been beneficial in a number of ways. It has given a very clear requirement for system developers and evaluators. It has had a positive influence in ensuring that experience gained from developing one system could be applied to another. It has facilitated the development of formal tools for assessing security, and so on. However there have been problems with systems based on the Bell and LaPadula model, not the least of these being the discovery, in many supposedly secure computer systems, of covert channels, that is means of communication which violate the security policy but were not foreseen in the security specification. In effect, the security specification was unable to support the stated security policy. This has not, hitherto, caused the foundation of the model to be challenged, but it has caused work to be undertaken on refining the model.

Since the original papers by Bell and LaPadula there have been a number of attempts to produce more general models which take into account system properties such as covert channels. Two well known examples are the non-interference model of Goguen and Meseguer [10] and Sutherland's work based on "possible world" semantics [18]. We can summarise both models by saying that they try to take into account information flow between two subjects no matter how it arises, whereas the Bell and LaPadula model is confined to constraints expressed in terms of components of the system state, such as files or directories. Nevertheless, despite the undoubted progress made by Goguen & Meseguer, Sutherland, and others, there still are considerable problems in building secure systems and verifying that they satisfy some stated security model. The aims of this paper are:

- to show that these problems are *inherent* in the nature of the models so far chosen;
- to show that the problems can only be overcome by choosing more appropriate bases for the models;

- to outline a more appropriate (enterprise-oriented) basis for security models.

1.2. The Characteristics of our Approach

We believe that a fundamental problem with these information flow models is that they take a "machine-oriented" view whereas we believe that it is only possible to articulate the meaning of security in the context of a computer system by including its operational and threat environment in the security model. Much of our paper is devoted to justifying this view, but it is helpful to draw out some of the issues which arise when one takes this approach to security.

First, we can give a simple definition of security:

> "A system is secure if it prevents *individuals* from achieving unauthorised access to information".[†]

The stress on individuals is essential – there are currently no laws that state that it is an offence for unauthorised computers or other machines to store or process classified information. More pragmatically we are only concerned with access of information by individuals as it is individuals who are responsible for initiating actions based on the information, and who are liable for the consequences of those actions.

Second, it is important to stress that what we have stated is the *fundamental* requirement – other issues such as control over information flow, auditing, etc. are *consequential*. They derive from a deeper analysis of the meaning of security concepts, from a consideration of implementation issues, and from recognition of the limitations of implementations. We will elaborate on these issues later in the paper.

Third, by being more precise about what we mean by a system, we can clarify our understanding, and improve the definition, of security. By a "system", we mean to imply that the entity is motivated by purpose, can make value judgements, respond to environmental change, and so on. Systems do not merely exhibit behaviour: it is that we ascribe meaning and purpose to that behaviour. In this context a computer is *not* a system, since we cannot answer questions concerning the security of its behaviour, i.e. we can't tell if display or transmission of information constitutes a breach of security.

In a system without a computer the interpretation of the above definition of security is clear:

- information is stored in documents;

[†] In general we should say "that which is of value", not information. In the context of military security it is normally assumed that information is the valuable resource to be protected, although as we shall see later this is not always the case

- security is achieved by enforcing constraints on, and monitoring, human access to documents, via a registry, security containers, etc. In other words access to information is controlled via access to data.

In a system with a computer we also must consider access to information processed by the computer. Thus:

- information also includes displays on screens, printouts, messages passed down communication channels, etc.;
- the definition of security does *not* refer to information inside the computer (since this cannot directly be accessed by people in the system).

This may seem to be an unnecessarily fine distinction as most stored information can be extracted from a computer, but it is an important point as it indicates that models based on machine concepts, even when discussed in apparently abstract terms such as subjects and objects, are focussing on information which is not significant from the point of view of security.

We are, nevertheless, in this paper principally concerned with *computerised* systems (CS), i.e. those which store and process some of their information on computers. A CS comprises: people (users and other individuals), computers and associated equipment, and documents and information made available at the computers' outputs. Borrowing the term from Checkland [6], we say that the computers reside in a Human Activity System (HAS). We can now provide an improved definition of security for a computerised system:

> "A CS is secure if it prevents unauthorised access by its users and other individuals both to its documents and to information input to, or output from, its computers."

However, this definition is still incorrect, albeit in a rather subtle way. We shall argue in Section 2 that security is related to the "flow of meaning" rather than the flow of information. For the moment though, we shall ignore the distinction until we give more justification for our belief that it is important. Whilst we will not rely heavily on this definition in the rest of the paper, it is helpful to include it as it indicates our basic reason for being interested in enterprise models. These models represent the HAS, including computer users and would-be security violators, thus enabling us to articulate security issues in a way which is consonant with the above definition.

1.3. Content of the Paper

In order to provide a justification for our proposed approach to security modelling, Section 2 gives an extensive account of the problems that arise in developing secure computer systems where the concept of security is based on the information flow approach. In particular, Section 2 endeavours to show that many of the problems which arise in supposedly secure systems are a direct *consequence* of the use of models based on the concept of information, and which do not encompass the HAS, or operational environment (context), for the computers. A specific problem which we stress is that the models relate to information, whereas they should relate to meaning.

Sections 3, 4 and 5 progressively build towards a new security model based on the concept of modelling the *enterprise* in which the CS resides. Section 3 discusses issues in the modelling of meaning via a brief discourse on formal semantics. Section 4 presents a new strategy for modelling meaning, and in Section 5 this strategy is used to outline an approach to enterprise modelling.

Section 2 concentrates on military security problems, since this is the domain in which the information flow models have been developed and used. The enterprise models are built up in a business context, because it is easier to explain the concepts in business terminology. However, the enterprise models are applicable in a military context, as we shall show in Section 6. Further, they enable us to take a more unified approach to security in the two domains recognising that military and commercial security are at different ends of a spectrum, rather than being antithetical.

In Section 6 we demonstrate how to apply the model in the context of database security, and also show by illustration how some of the standard military security issues can be treated in our approach. The presentation both describes some aspects of the new model, and indicates how it overcomes some of the problems identified in Section 2.

The work on this new model is by no means complete: Section 7 draws some conclusions from our work and indicates the direction of future research work on the model.

2. INFORMATION FLOW MODELS: FLAWS AND LIMITATIONS

2.1. Criteria for Assessing the Models

In order to be useful, security models have to be abstractions of reality which clarify the concepts and principles associated with developing secure systems. They also have to correlate sufficiently well with reality to be helpful and not misleading. We can think of failings to satisfy these criteria as being **synthetic**[†] flaws or limitations of the models.

Further there are some requirements on the models which are independent of their application. The most obvious of these is the requirement for consistency. We can think of these as being **analytic**[†] requirements, hence a failure to satisfy the requirements would be viewed as an analytical limitation.

The difference between synthetic and analytic faults is the classical one: an analytic fault in a model is a defect in the model itself which can be determined without enquiring into what the model is a model of; a synthetic

[†] We adopt these terms following a long philosophical tradition which includes Leibniz and Kant.

fault is one which does not invalidate the model *per se*, only its relation to the reality being modelled. A good model is one that has neither kind of fault. A characteristic of a model without analytic faults is that it is logically sound and made by a modeller who understands the nature of logic and mathematics; a characteristic of a model without synthetic faults is that it is fit for its purpose and made by a modeller who understands the nature of the problem being modelled.

It is a basic tenet of our work that information flow models (IFMs) do not have the above desirable characteristics and, indeed, that the models are themselves the cause of many of the problems associated with the production of secure computerised systems (SCS). So far as we are aware the main models do not suffer from major analytical faults. However, we believe that they are weak from a synthetic point of view. In order to justify this belief we shall consider a number of problems which arise when developing SCS using the IFMs. We then argue that the source of these problems is the way in which the IFMs use information theory as a basis. This is not to say that information theory is inappropriate or flawed, but that the models make an inappropriate association between the flow of information and the "real world" meaning of security (or, perhaps more accurately, security violations).

We have grouped the problems in to related classes, the first three of which represent divergences between the "real world" meaning of security and that defined by the IFMs. The fourth group deals with technical modelling problems to do with reference, that is, with the relationships between entities in the model and entities in the real world. The order of presentation of the problems is not especially significant; indeed, they can be thought of as different facets of the same problem, rather than essentially different problems. Nonetheless the first group of problems is undoubtedly the crux of the matter.

We do not pretend that the list of problems is exhaustive. However, we hope that the list is sufficiently extensive, and that the nature of the problems is sufficiently fundamental, to make it clear that a replacement model is required, and that merely applying cosmetic changes to existing models will be ineffective.

2.2. Information, Meaning and Data

A key issue in our criticism of the the IFMs stems from the distinction between information, meaning and data; thus it is important to understand their relationships (at least as we use the terms).

Data is the carrier for information (although this use of the term means that we may have to treat unusual items, e.g. electrons, or puffs of smoke from the Vatican chimney, as data). We use information in the sense of information theory -- information is deemed to be conveyed by some data where it reduces some set of possibilities, in other words data conveys information if the recipient "knows more" about some situation as a result of receiving the data. Meaning is the interpretation we make of some information in a *context*. Essentially the context is an intellectual one, e.g. the human knowledge necessary to understand natural language. However, context may also be amplified or aided by mechanical means. Examples of aids are information stored in a machine, e.g. a decryption algorithm and key, or a database which enables associations to be made from some input information.

By way of example consider the receipt of the character string "Bush is the new president-elect". The data is the character string. There is information associated with this data as the number of possibilities for president-elect has been reduced from two to one. There is also a (much richer) meaning associated with this information in terms of the effect on US foreign policy, the effect on balance of trade deficit, and so on.

It should be noted that the distinction between information and meaning is somewhat subtle, and it is arguable whether or not information theory is concerned with meaning (see section 2.7). However, in the following discussion we work on the basis that information theory is genuinely concerned with meaning, but the IFMs are concerned with information in the more limited sense illustrated above.

2.3. Semantic Problems

The problems presented in this section are related in that they all concern distinctions between information, as employed by the IFMs, and meaning as described above. In particular they reflect the fact that "real world" security problems are based on the *meaning* of data, whereas the models only predict effects (e.g. security breaches) which are based on the information carried by the data. This leads to a considerable divergence between the models and the true concerns of security.

2.3.1. The Meaning Problem

In general it is not possible to tell if a security breach has occurred just from knowledge that there has been passage of information (to an individual).

- For example, if someone receives a file which they cannot decrypt then information has flowed but security has not necessarily been breached (assuming that knowledge of data volumes, and knowledge of the data's existence are not significant). A judgement regarding security can only be made in terms of the operational environment for the system.
- Let us make the reasonable assumption that the CPU identifier of some common service machine is unclassified, and is known to all processes. If the CPU identifier is sent down a covert channel from a high to a low subject, then the IFMs would indicate that a flow, and hence a security violation, had occurred. In fact security has not been breached as the

recipient has not learnt anything new (though in fact the situation is slightly more complex, see section 2.4.2).

In the first case the problem is that the IFMs do not allow for the effect of the processing of the information. In an accurate interpretation of information theory, the requisite distinction can be made as the conditional entropy (information content) of the message when it reaches the human recipient depends upon whether or not the recipient can decrypt (possibly with computational assistance) the data. The IFMs do not make this distinction as they ignore the effects of the processing of the message, and the (significance of) other information available to the recipient. Another way of looking at this is that there is more than one possible interpretation of the information depending on the availability of the encryption key, of which some violate security and the some do not – but the IFMs don't make the distinction.

In the second case the IFMs are really describing the data flow and not the information flow, nor the meaning associated with the data flow. Again this is because of lack of a model of context.

Thus IFMs are misleading in that they may suggest that security breaches have occurred where they have not. The key point is that we cannot make judgements about security unless we know what the information *means* to the recipient. This point is *fundamental*, since IFMs do not give us a basis for modelling meaning (see Section 2.7). Pragmatically we may say that the models are too **strong** in that they deny behaviour which is in fact secure.[†]

2.3.2. The Aggregation Problem

Information flow does not account for aggregation; that is, the fact that the composition of objects at a low level of classification can produce one object at a much higher level of classification.

- For example, the conjunction of an unclassified name, latitude/longitude pair, and a pairing of date and time might indicate the whereabouts of a submarine, in a suitable *context*. Thus conjoining three unclassified items can produce a highly classified result.

Again the issue is meaning: the meaning of the composite object is not simply the "sum" of the meanings of the parts. Thus, in this case, the IFMs are too weak – that is, they permit behaviour which contravenes security.

[†] In general one will be willing to accept some restrictions on secure behaviour in order to prevent insecure behaviour. This one point alone is not enough to shake the foundations of the IFMs.

2.3.3. The Inference Problem

Inference is a process where some (sensitive) information is deduced from other (less sensitive) information; this process is not predicted or modelled by the IFMs.

- For example, it is possible to derive classified information from (statistical) databases even where the responses to individual queries are restricted to unclassified data.

This is similar to the aggregation problem; indeed, one could view inference as being a situation where aggregation occurs in the mind of the recipient, and it shows another area in which the models are too weak.

2.3.4. The Reductionist Problem

The IFMs apply to all "units" of information, thus falsely attributing classifications to meaningless items such as the words "the", "and" and "but" in isolation.

- For example, editing a Top Secret paragraph, removing the words, can lead to having a Top Secret full-stop, as the classification stays attached to the remnants of the paragraph.

This problem arises from what we might term the reductionist† fallacy, that properties of the whole apply also to its parts. The extreme case illustrated above can, of course, easily be handled, but it is a subtle semantic problem to decide exactly when the edited paragraph can be downgraded. Again the fundamental issue is the inability of the IFMs to model meaning, although in this case we could argue that they do not even model information accurately.

Thus, in this case, the IFMs are too strong as they erroneously attribute sensitivity (high levels of classification) to innocuous information.

2.3.5. Summary

The above examples illustrate that the IFMs do not reflect "real world" security issues because they are based on the concept of information not meaning. If the IFMs were consistently conservative, that is they caught all potential security breaches, but did not overconstrain valid behaviour then they might be acceptable. However, the IFMs are both too strong and too weak and so they are not acceptable by this argument.

Over the years a number of attempts have been made to adapt the IFMs to overcome specific problems, e.g. inference and aggregation, but we know of no systematic attempt to address the complete range of problems we have identified above. So far this work seems only to be capable of making

† The term again reflects philosophical tradition, this time stemming from Descartes.

minor incremental improvements to the models to deal with specific problems. We believe that the limited nature of such success is inevitable — intuitively, it is very difficult to make the models simultaneously both stronger and weaker. As we shall see later, we need a much richer model of context even to begin to treat these problems in a systematic way.

2.4. Functionality Problems

There are various problems with the models which constrain or adversely affect the functionality which it is possible to implement in a computer whilst ensuring that the computerised system is secure. It is necessary that there should be some constraints — in a sense that is why we have the models — but some of the constraints conflict with basic operational requirements.

We deal first with two problems which relate to the so-called covert channels, then present two problems which stem from the fundamental requirement to process information of more than one classification in the same system.

2.4.1. The Covert Channel Problem

A notorious problem with supposedly secure computer systems is that they can contain so-called covert channels. These covert channels are means of communicating information between subjects by indirect means, e.g. by modulating system resources, which could contravene security. Even given our view of security, covert channels are an issue as it is possible to use them to make sensitive information available at the outputs from computers.

The Bell and LaPadula approach does not cater for covert channels. The IFMs were developed at least partially to deal with this problem; however, they introduce a further problem, which we can best illustrate by means of the following discussion.

By the information flow definition, "secure" means "no illegal information flow", that is no information flows from high subjects or objects to low ones. However, there is a consequence of this definition:

- A system is either insecure or has impaired functionality or performance.

We can show that this is a consequence by means of the following informal reasoning. In order to provide functionality it is necessary to use resources, e.g. CPU time and storage. In general information can be passed by modulating resources, and attempts to block channels tend only to move them elsewhere, or to change their nature, e.g. from storage to timing channels. Information flow can only be controlled (eliminated) by constraining resource usage, i.e. by impairing functionality or performance.

We can clarify this general principle by means of a simple example. Consider a one way regulator, that is a device designed to allow classified information to flow in one direction (from low to high) but to deny flow in the reverse direction. For such a regulator implemented using handshaking protocols over wires or a bus there will inevitably be reverse information flow through the protocol messages. For a regulator using an optical isolator (i.e. one that transmits photons in only one direction) there is no reverse flow, but there is a chance that information will be lost because of synchronisation problems and possibly buffer overflow, i.e. the functionality is modified (impaired).

Thus systems either have impaired functionality, covert channels, or both (but see the discussion on bandwidth below). In the limit, to prevent information flow by covert channels, we have to constrain systems to having no functionality!

2.4.2. The Bandwidth Problem

In practice security is not an absolute concept, but a relative one, and we are concerned about the bandwidth of information flow. This is a guiding principle in encryption where we accept a cipher as adequately secure if the time and cost of cryptanalysis are large with respect to the longevity (currency) and value of the data. In other words the bandwidth of data to unauthorised recipients of data should be acceptably low.

This general observation about bandwidth is very significant in the context of covert channels.

- If a covert channel exists, but it only has a bandwidth of 1 bit per day, then security standards (e.g. the Orange Book) say that the channel can be ignored (although the IFMs would still treat this as an illegal flow if it was from a high to a low subject or object). However, if the single bit *means* "the nuclear strike is on tomorrow" then (presumably!) a security breach has occurred despite the low bandwidth. Clearly the meaning in this case would have had to be agreed by the sender and recipient previously, and outside the computer system.

What we have identified here is a "low bandwidth, high meaning" channel. It should be noted that our example in section 2.3.1 of passing the CPU identifier could provide a single bit channel (send or don't send) which shows one of the relationships between these examples. Arguably it is therefore a strength of the IFMs that they do cover all forms of flow, independent of the bandwidth, but in the above example the single bit flow could be disastrous even if it did not contravene the flow rules, e.g. it went from a low to a high subject. Again the basic problem is the absence of a model of context for interpreting the data (or information) and assigning a meaning to it.