A Review of Security Models and Information Flow

Kevin O'Neill

March 6, 2001

In his 1990 paper [1], McLean tries to pin down a theory of information flow, developing a flow-based security model called FM which he compares with other theoretical models of security including Nondeducibility, Noninterference, Generalized Noninterference, and the Bell and LaPadula access control model.

1 Nondeducibility

The Nondeducibility model presented by Sutherland [2] tries to formalize what it means for information to flow from high-level objects to low-level objects, but the formalization doesn't allow information to low from high to low, *or* from low to high—it's better suited to compartmentalization that to security.

Formally, given H, an assignment sequence to a high-level input port, and L, an assignment sequence to low-level input *and* output ports, Nondeducibility states that information flows from high to low if:

$$p(H) > 0 \wedge p(L) > 0 \Rightarrow p(H \mid L) > 0$$

or, equivalently,

$$p(H \mid L) = 0 \Rightarrow p(H) = 0 \vee p(L) = 0$$

Less formally, nondeducibility says that we cannot use low-level information of any kind to rule out any possible high-level input. This definition has several weaknesses:

- Nondeducibility is symmetric in terms if H and L. Given high-level input, high-level users are not able to rule out any any low-level sequences, which is clearly undesirable.

- Nondeducibility is too strict. It rules out event sequences which aren't insecure. As an example, consider a sequence where a low-level user gives some arbitrary number as input which is echoed to the high-level user, who must then enter some *different* number. The low-level user can "rule out" any high-level input sequence where the high-level user chooses the same number as the low-level user because of the high-level protocol; however, it's clear that whatever we want our definition of "information flow" to be, information does not flow from high-level users to low-level users in this scenario.

- Just because a low-level user is unable to *rule out* any high-level sequence doesn't mean she is unable to determine a probability distribution on high-level sequences. If one high-level sequence is very likely based on low-level input/output sequences while all other sequences are very unlikely, the system is clearly not secure; we would like to be able to say that information does flow to low-level users in this scenario.

One solution that addresses all of these concerns is to introduce some form of time or state into the model. If we let L_t and H_t refer to assignment sequences in some state t, we can prohibit information flow from H to L by requiring that $p(L_t \mid H_{t-1}, L_{t-1}) = p(L_t \mid L_{t-1})$, i.e., that low-level output in some state t is independent of the high-level input in the previous state $t-1$. (Conditioning on L_{t-1} ensures that the low-level user cannot combine low-level knowledge from states t and $t-1$ to learn something about high-level knowledge at state $t-1$.)

This model prohibits information flow from H to L, but not from L to H; in addition, it rules out systems in which low-level users can gain *probabilistic* knowledge about high-level events. The problem is that this model is still too strict. L_t may be statistically correlated with H_{t-1} without being causally influenced by it. This could happen, for example, if the values of both were influenced by some previous low-level event. It should be clear that without causality, statistical correlation shouldn't be a security problem.

2 McLean's Flow Model

In an attempt to limit the notion of information flow to cases where high-level input has some sort of causal influence on low-level outputs, McLean considers the sequence of value assignments $\mathbf{H}_t$ (and $\mathbf{L}_t$), which represents the values assumed by high-level (and low-level) objects in every state preceding state t.

McLean's *Flow Model* (FM) states that a system is secure only if $p(L_t \mid \mathbf{H}_t, \mathbf{L}_t) = p(L_t \mid \mathbf{L}_t)$, i.e., that the low-level input at time t is independent of all previous high-level input, given all previous low-level input. Under this model, a statistical correlation between high-level and low-level inputs would be fine if both were influenced by some previous low-level event: conditioning on $\mathbf{L}_t$ guarantees that L_t and $\mathbf{H}_t$ are indeed independent.

It's clear that any model of information flow must address all security-relevant causality considerations, and this may be an area for future research.

References

[1] John McLean. Security models and information flow, 1990.

[2] David Sutherland. A model of information. In *Proc. 9th National Security Conference*, pages 175–183, Gaithersburg, Md., 1986.