An Entropy Conservation Law for Testing the Completeness of Covert Channel Analysis

Randy Browne
147 Old Bloomfield Avenue
Parsippany, New Jersey 07054
Phone: (201) 244-0612
E-mail: randybrowne@delphi.com

Abstract

Covert channel analysis typically involves study of individual covert channels in isolation, and determining the thoroughness of such case-by-case analysis can be difficult. To help address this problem, this paper formally defines the notion of a "complete" set of covert channels. Informally, a set of covert channels is "complete" when those channels in the set can operate in tandem to produce the maximum possible covert information flow out of a system. More formally, a "complete" set of covert channels is defined as a solution to an equation called the Maximum Information Flow Equation. An alternate way of expressing "completeness" for sets of covert channels is that all "complete" covert channel sets, and only "complete" sets, always satisfy a certain Entropy Conservation Law, which is given in different forms. One form of the Entropy Conservation Law is that any "complete" set of covert channels can be used to represent overall system behavior in what the author calls Covert Channel Normal Form. Although this paper is mainly theoretical in nature, the author also discusses some possible ways of using the theory, along with open issues.

1 Introduction

Any multilevel system that has any covert channels at all, can be regarded as having one single complex covert channel that induces some overall, global covert information flow. From the perspective of information theory [Shannon48], calculating this global information flow *could* be viewed as the *ideal* objective of covert channel analysis. But in practice, covert channel analysis is usually limited to the study only of individual covert channels (using a variety of analysis techniques [NCSC93]) without explicit consideration of the global covert information flow.

CCS '94- 11/94 Fairfax Va., USA
© 1994 ACM 0-89791-732-4/94/0011..$3.50

However, such case-by-case analysis of individual, isolated covert channels invites questions about the thoroughness (completeness) of the analysis. Clearly, the combined effect (information flow) of any set of individual covert channels cannot exceed the global covert information flow (whatever that global flow is), but if we wish to capture the global covert information flow merely by studying individual covert channels, there are two problems we must deal with:

1. The *scope* problem: In order to compute the global covert information flow, do we have to find all covert channels in a system? Or is there some subset of all possible covert channels that we can use to capture the global covert information flow in some way and, if so, how do we define such a covert channel subset?

2. The *quantification* problem: Given a set of covert channels that is supposed to capture the global covert information flow, how can we actually compute or estimate the global (aggregate) covert information flow from the channels in the set?

Although interrelated, the second problem (quantification) is *not* the subject of this paper. Information theory offers at least a partial answer to the quantification problem of calculating the aggregate information flow [Shannon48, Shannon61]. Rather, the main topic of this paper is the scope problem; identifying enough covert channels so that the global covert information flow is fully captured by those channels.

To actually capture the global covert information flow for a system, it turns out that it is not generally necessary to discover all covert channels as such; only some subset (in general) which I will formally define later as a *complete* set of covert channels. Complete sets of covert channels are analogous to the basis of a vector space [Halmos74] in that, the covert information flow from the combined action of a complete set of covert channels, is the same as for any other complete set, and also the same as for the system's overall global covert information flow.

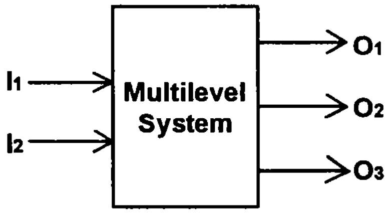
After discussing a simple example system which illustrates both the scope problem and the quantification problem, I will give a simple formal theory of covert channels as a basis for defining what is meant by a complete set of covert channels. After defining the concept of completeness for sets of covert channels, I will show that the notion of completeness is equivalent to whether or not a given set of covert channels "conserves entropy" in the sense defined by a certain *Entropy Conservation Law*. The Entropy Conservation Law is presented in different, but equivalent forms. One form in particular is the idea of describing a multilevel system's behavior in what is called *Covert Channel Normal Form* (CCNF). I will show that if one can succeed in expressing a multilevel system's behavior in CCNF using *only* a given set of covert channels, this is a necessary and sufficient guarantee that the given set of covert channels is complete.

Finally, I discuss some potential ways of using the notion of complete sets of covert channels in security engineering. However, instead of using the Entropy Conservation Law (and the attendant notions of CCNF and completeness of covert channels) to solve the *general* problem of covert channel analysis, I will rather suggest that it is better to focus on *specific* security architectures that lend themselves easily to CCNF and the related notions. I will discuss an approach to (selectively) applying the formal theory and various open issues after the theory is set out.

2 Global versus Local Information Flow

To illustrate the difficulties in determining global covert information flow from a collection of isolated covert channels (so-called "local information flow"), we start with a simple "toy" problem that illustrates how a covert channel analyst might fail to detect a "complex" covert channel with non-zero capacity from a study of suspected covert channels, each of which has been found to have zero capacity.

Consider a simple multilevel system with two binary SECRET (or "high") input channels, and three binary UNCLASSIFIED (or "low") output channels:



We assume that the system contains an internal clock and upon each clock 'tick', the pair of inputs are sampled and the three outputs are generated according to the following four probability tables (showing the conditional probability of each 3-bit binary UNCLASSIFIED output pattern for each 2-bit binary SECRET input pattern):

I_1=0 & I_2=0	O_1	O_2	O_3	Probability
	0	0	0	3 / 56
	0	0	1	11 / 56
	0	1	0	11 / 56
	0	1	1	3 / 56
	1	0	0	11 / 56
	1	0	1	3 / 56
	1	1	0	3 / 56
	1	1	1	11 / 56

I_1=0 & I_2=1	O_1	O_2	O_3	Probability
	0	0	0	13 / 72
	0	0	1	5 / 72
	0	1	0	5 / 72
	0	1	1	13 / 72
	1	0	0	5 / 72
	1	0	1	13 / 72
	1	1	0	13 / 72
	1	1	1	5 / 72

I_1=1 & I_2=0	O_1	O_2	O_3	Probability
	0	0	0	1 / 16
	0	0	1	3 / 16
	0	1	0	3 / 16
	0	1	1	1 / 16
	1	0	0	3 / 16
	1	0	1	1 / 16
	1	1	0	1 / 16
	1	1	1	3 / 16

I_1=1 & I_2=1	O_1	O_2	O_3	Probability
	0	0	0	1 / 32
	0	0	1	7 / 32
	0	1	0	7 / 32
	0	1	1	1 / 32
	1	0	0	7 / 32
	1	0	1	1 / 32
	1	1	0	1 / 32
	1	1	1	7 / 32

In discussing this example, I will use the notation such as $\{I_1,I_2\} \Rightarrow \{O_3\}$ to denote the covert channel where both SECRET inputs are controlled but *only* the third UNCLASSIFIED output is monitored as a response. I make the following claims that can be verified by examining the above tables:

1. The "system-wide" covert channel involving all inputs and outputs $\{I_1,I_2\} \Rightarrow \{O_1,O_2,O_3\}$ has *non-zero* capacity, because, if the inputs are held steady for an extended period of time, large samples of output triples will tend to have a relative frequency that matches the unique output distribution (shown by one of the four tables above) for the given fixed pair of inputs.

2. Every covert channel that monitors fewer than three outputs has *zero* capacity, such as $\{I_1\} \Rightarrow \{O_2\}$, and even covert channels such as $\{I_1, I_2\} \Rightarrow \{O_2, O_3\}$ using as many as two outputs, etc.. Indeed, channels that have only a single (binary) output will produce a 0 or 1 with probability 1/2 *regardless* of the input, and similarly, channels with only two outputs will produce each of the four possible output combinations with probability 1/4 also *regardless* of the input.

Now let us consider two covert channel analysts (A and B) who have been assigned to assess the covert channels in our system. We will suppose for the sake of illustration that, for whatever reason, the analysis of the "system-wide" covert channel $\{I_1, I_2\} \Rightarrow \{O_1, O_2, O_3\}$ is too complex to be analyzed directly, so our analysts want to try to discover this channel via an analysis of combinations of simpler channels such as $\{I_1, I_2\} \Rightarrow \{O_3\}$, $\{I_2\} \Rightarrow \{O_1\}$, etc. Now suppose our two analysts discover and analyze various covert channels (every one of which has zero capacity) as shown here:

Analyst A:

$\{I_1\} \Rightarrow \{O_1\}$
$\{I_2\} \Rightarrow \{O_1\}$
$\{I_1\} \Rightarrow \{O_3\}$
$\{I_2\} \Rightarrow \{O_3\}$
$\{I_1, I_2\} \Rightarrow \{O_1\}$
$\{I_1, I_2\} \Rightarrow \{O_3\}$
$\{I_1, I_2\} \Rightarrow \{O_1, O_3\}$

Analyst B:

$\{I_1\} \Rightarrow \{O_2\}$
$\{I_2\} \Rightarrow \{O_1\}$
$\{I_1\} \Rightarrow \{O_3\}$
$\{I_2\} \Rightarrow \{O_3\}$

Now one might wonder which of the two analysts has done a more thorough job of covert channel analysis? Since both analysts have produced lists of zero capacity channels and neither analyst has detected the "system-wide" covert channel $\{I_1, I_2\} \Rightarrow \{O_1, O_2, O_3\}$, is this even a meaningful question?

As I will show later, this is indeed a meaningful question. The fact that the overall channel capacity is non-zero whereas the known channel capacities are zero is the *quantification* problem; not the *scope* problem. Using the Entropy Conservation Law (to be given later), we can still ask and get an answer to the question about whether our analysts have identified enough channels (the scope problem) in order to calculate the overall covert channel information flow (the quantification problem).

The answer as to which if either of our analysts has done a thorough job of covert channel analysis may seem surprising at first, but will seem obvious after seeing the formalism. The fact is, Analyst B's much shorter list is a complete set of covert channels in the sense I will define; Analyst A's longer list containing more complex channels involving multiple outputs is, for all his/her effort, incomplete. We will return to this later, but a hint as to why Analyst A's list is incomplete is that, note that Analyst A has no channels involving the second UNCLASSIFIED output O_2 which is one of the outputs in the "system-wide" channel $\{I_1, I_2\} \Rightarrow \{O_1, O_2, O_3\}$. This will turn out to be a special case of the Entropy Conservation Law. It isn't necessary in general to find channels that involve *all* system outputs; just those outputs which might contain covert channels (which I will formalize later and which happens to be *all* system outputs in our "toy" problem).

3 A Simple Covert Channel Theory

To define the notion of completeness for covert channels, we must first have a formalism for discussing covert channels in general. In this section I give a simple covert channel theory adequate for our purposes. Except for the formal definition of a covert channel (below), the model described in this section is essentially nothing more than a notational variant of the Turing Test Model [Browne91, Browne94].

3.1 A Closed Multilevel System Model

As with the Turing Test Model [Browne91, Browne94], we deal only with multilevel state machines that are closed ("input-free", "self-running", "self-driving", or "perpetual"). The restriction to closed machines is no loss of generality since, one can always consider a sequence of external inputs as being encoded in the initial state of the machine, either by explicitly encoding input sequences, or encoding input implicitly in "wound-up" form such as by defining an "input strategy" [WittboldJohnson90] or a "generator process" [Browne91]. I model a system as a collection of random variables and functions defined as follows, where parameter t corresponds to (discrete) time, and ℓ is a security level:

1. The *information variable* $\mathcal{V}$[1] can be thought of as a summary of all information being protected by a security policy. So that there is no loss of generality in power expressing security policies, we suppose that $\mathcal{V}$ is the initial state of some particular (closed) universal machine. $\mathcal{V}$ encodes the state transition relation for some particular machine to be "emulated" as well as the input and initial state information for the "emulated" machine at all different security levels. $\mathcal{V}$ has no specific security level, but different "views" of $\mathcal{V}$ (below) do have an associated security level.

2. The (deterministic) function $View(\mathcal{V}, \ell)$ extracts the information about $\mathcal{V}$ known to security level ℓ, such as

[1] Since the covert channel model I am describing here is mainly just a notational change from the Turing Test Model [Browne91, Browne94], I am only partly describing the model. The reader interested in a full description should refer to [Browne91, Browne94]. The covert channel model being described is actually a multi-player game where the information variable $\mathcal{V}$ (called the "Turing Test Variable" in [Browne91, Browne94]), can be chosen from a variety of probability distributions. The nature of this game is not very important for understanding the central points of this current paper. However, the fact that $\mathcal{V}$ can be chosen from a variety of distributions is partly relevant to understanding the notation used in the Maximum Information Flow Equation, given later.

a particular machine's state transition relation (always known to all security levels; open design), the input strategies of all levels at or below level ℓ, and the initial state variables at or below security level ℓ for the machine designated by $\mathcal{V}$.

3. The (random) function $Output(\mathcal{V},\ell,t)$ indicates, for security level ℓ, the output produced by a particular "run" or "random trial execution" of the machine denote by $\mathcal{V}$, when driven by the various "input strategies", initial state, etc. as indicated (also) by $\mathcal{V}$. (We assume that $View(\mathcal{V},\ell)$ determines the value of $Output(\mathcal{V},\ell,0)$.)

3.2 A Formal Model of Covert Channels

We also add to our basic definition of a multilevel state machine, various auxiliary random functions that represent covert channels in the system. A *covert channel* is any time-dependent random function written as $C(\mathcal{V},\ell,t)$ with a random value that is determined by the system output, meaning:

$$\forall\ell\,\exists h\,\forall t\,\forall\mathcal{V} \quad h(Output(\mathcal{V},\ell,t)) \;=\; C(\mathcal{V},\ell,t)$$

where h is any function from the range of $Output(\mathcal{V},\ell,t)$ to the range of $C(\mathcal{V},\ell,t)$. The only important attribute of function h is that h depends only on ℓ and on $C(\mathcal{V},\ell,t)$ so that every covert channel is essentially a property of the system output (as seen at some security level ℓ).

3.3 Ensemble Covert Channels

If $C_1,\ldots.,C_n$ is a list of covert channels, the *ensemble* of those channels is that random function $C(\mathcal{V},\ell,t)$ such that:

$$C(\mathcal{V},\ell,t) \;=\; <x_1,\ldots.,x_n>$$

where $\forall k\; C_k(\mathcal{V},\ell,t) = x_k$. I remark that by this definition, the ensemble for any set (list) of covert channels $C_1,\ldots.,C_n$ is *itself* a covert channel as formally defined; meaning that the ensemble is *also* determined by $Output(\mathcal{V},\ell,t)$ whenever all of $C_1,\ldots.,C_n$ are so determined.

3.4 The Maximum Information Flow Equation

The Maximum Information Flow Equation given below is essentially the same as the "Generalized Turing Test Condition" [Browne94], except for notational changes and renaming of variables. The Equation can be thought of as an abstract and probabilistic form of Conditional Non-Interference [GoguenMeseguer84] and relates probabilities of the information variable $\mathcal{V}$ given various other kinds of information. Formally, I say that a covert channel $C(\mathcal{V},\ell,t)$ has *maximum information flow* if and only if the *Maximum Information Flow Equation* (here) is satisfied[2] for $C(\mathcal{V},\ell,t)$:

$$\begin{aligned}&\forall w\,\forall x\,\forall y\,\forall z\,\forall\ell\,\forall t\\ &\quad pr(\mathcal{V}=w \mid Output(\mathcal{V},\ell,t)=x,\\ &\qquad C(\mathcal{V},\ell,t)=y, View(\mathcal{V},\ell)=z)\\ &\quad = \; pr(\mathcal{V}=w \mid C(\mathcal{V},\ell,t)=y, View(\mathcal{V},\ell)=z)\end{aligned}$$

3.5 Complete Sets of Covert Channels

Next, I say that a set (list) of covert channels $C_1,\ldots.,C_n$ is *complete* if and only if their ensemble covert channel $C(\mathcal{V},\ell,t)$ has maximum information flow (that is, satisfies the Maximum Information Flow Equation). The sense in which the Maximum Information Flow Equation really maximizes covert information flow is expressed by the following theorem:

Theorem I. Consider n covert channels $C_1,\ldots.,C_n$ and let $\Delta H(\mathcal{V}\,|\,out)$ be the average entropy change[3] in $\mathcal{V}$ given $Output(\mathcal{V},\ell,t)$. Also, let the term $\Delta H(\mathcal{V}\,|\,C)$ be the average entropy change in $\mathcal{V}$ given knowledge of the covert channels $C_1,\ldots.,C_n$. Then the covert channels $C_1,\ldots.,C_n$ are complete if and only if:

$$\Delta H(\mathcal{V}\,|\,C) \;=\; \Delta H(\mathcal{V}\,|\,out)$$

Proof: By Theorem A of [Browne94], if $C_1,\ldots.,C_n$ are complete, then we have that:

$$\Delta H(\mathcal{V}\,|\,C) \;\geq\; \Delta H(\mathcal{V}\,|\,out)$$

But because the system output determines the values of the covert channels, we can conclude that:

$$\Delta H(\mathcal{V}\,|\,C) \;=\; \Delta H(\mathcal{V}\,|\,out)$$

Conversely, assume the preceding holds. Now, by the definition of a covert channel, we have:

$$\Delta H(\mathcal{V}\,|\,out,C) \;=\; \Delta H(\mathcal{V}\,|\,out)$$

and the two preceding assertions imply that:

$$\Delta H(\mathcal{V}\,|\,out,C) \;=\; \Delta H(\mathcal{V}\,|\,C)$$

From the preceding, the ensemble channel for $C_1,\ldots.,C_n$ must be a solution to the Maximum Information Flow Equation and thus, by definition, the channels $C_1,\ldots.,C_n$ are complete.
□

[2] With regard to my previous remark that random variable $\mathcal{V}$ can be chosen from a variety of probability distributions, the Maximum Information Flow Equation must hold regardless of how $\mathcal{V}$ is chosen. Consequently, the reader should view the term "*pr*", which denotes probability in the Equation, as a free variable ranging over various probability distributions corresponding to different ways that $\mathcal{V}$ may be chosen; for details, see [Browne91].

[3] The average entropy change is also known as the "mutual shared information" between the random variables involved.

The preceding theorem justifies a loose analogy with the basis of a vector space [Halmos74][4] in that the information flow of *all* complete sets of covert channels is the same. Furthermore, the covert information flow from any/all complete sets of covert channels achieves the maximum, global covert information flow occurring through the entire system output.

Also, an absolutely crucial property of complete sets is that they are "stable"; meaning that any extension of a complete set of covert channels is complete. While this may seem a trivial comment, it is not as obvious as it first sounds. The example presented earlier involving our two Analysts A and B has already shown one case where adding random variables to a model can strictly increase information flow. The "stability" of the completeness notion is paramount to sensibility of this theory; otherwise, one might "stumble" across a "complete" set of channels only to find that by adding additional channels, information flow increases, making the notion of "completeness" absolutely worthless.

Theorem II. If S is a complete set of covert channels, and T is any set of covert channels, then $S \cup T$ is complete.

Proof: Since S is complete, then by Theorem I:

$$\Delta H(\mathcal{V}\,|S) \;=\; \Delta H(\mathcal{V}\,|out)$$

And obviously:

$$\Delta H(\mathcal{V}\,|S,T) \;\le\; \Delta H(\mathcal{V}\,|S)$$

but since $S \cup T$ is a set of covert channels (not yet claimed to be complete), we have by the definition of a covert channel that:

$$\Delta H(\mathcal{V}\,|out) \;\le\; \Delta H(\mathcal{V}\,|S,T)$$

whereby all of the preceding prove:

$$\Delta H(\mathcal{V}\,|S,T) \;=\; \Delta H(\mathcal{V}\,|out)$$

And by using Theorem I in the other direction, we have finally that $S \cup T$ is complete.
□

4 An Entropy Conservation Law

We now come to a simple but general Entropy Conservation Law that, in various forms, captures the concept of completeness for covert channels. The Entropy Conservation Law can be viewed as an extension of the Separability Principle put forth by [Rushby81]. Roughly speaking, the Separability Principle requires that "the system output observed by an entity must be entirely explained by that entity's input to the system". The Entropy Conservation Law extends this by saying, loosely, that "if system output is not determined by an entity's input, then, whatever is 'left over' must be entirely explained by covert channels and real noise".

4.1 Abstract Entropy Conservation Law

We start with an informal statement of the Entropy Conservation Law:

Informal Statement [First Form of the Entropy Conservation Law]. For any (closed) multilevel information system, if an onlooker with clearance ℓ is "fully informed"; meaning, is aware of all information authorized at level ℓ, and may even have some information not authorized at level ℓ, then whatever the onlooker expects of the variability in the system output ***must always be due entirely*** to the anticipated variability of covert channels, plus possibly some noise source. That is, for a "fully informed onlooker" of a closed multilevel system:

Output Uncertainty =
Covert Channel Uncertainty + *Noise*

For illustration sake, two special cases of the Entropy Conservation Law are, first, if a (closed) system has no covert channels, then whatever uncertainty one has in the system output must be due entirely to system noise (assuming full awareness of all authorized knowledge; input, initial state information, etc.). Second, if a (closed) system is deterministic, then whatever uncertainty one has in the system output must be due entirely to unpredictable behavior in the system's covert channels.

Further insight into the Entropy Conservation Law is possible with a more formal statement; in particular, more can be said about the noise source than is evident from the informal statement, and also, the informal statement doesn't indicate what "entropy conservation" has to do with completeness for sets of covert channels. However, to formalize the Entropy Conservation Law, we need to formalize what we mean by a "fully informed onlooker". I model a *fully informed onlooker with security clearance* ℓ as a random process $\mathcal{X}(\mathcal{V},\ell,t)$ such that:

$$\begin{array}{l}
\exists f \;\; \exists g \;\; \exists h \;\; \forall t \ge 0 \\
\quad f(\mathcal{V}) = \mathcal{X}(\mathcal{V},\ell,0) \quad \wedge \\
\quad g(\mathcal{X}(\mathcal{V},\ell,0)) = View(\mathcal{V},\ell) \quad \wedge \\
\quad h(\mathcal{X}(\mathcal{V},\ell,t)) \; : \; \text{one}-\text{one} \quad \Rightarrow \\
\qquad\qquad\qquad < Output(\mathcal{V},\ell,t), \mathcal{X}(\mathcal{V},\ell,0) >
\end{array}$$

This cryptic expression has a simple explanation. The first term above involving function f means that, in our closed system world, everything that "$\mathcal{X}$ knows initially" is determined by the information variable $\mathcal{V}$; there is nothing else to know about initially, except about $\mathcal{V}$, so there is some function f that extracts from $\mathcal{V}$ what initially determines what "$\mathcal{X}$ knows". Further, function g indicates that whatever "$\mathcal{X}$ knows initially" includes *all* knowledge of $View(\mathcal{V},\ell)$, so the existence of some function g indicates that there is some way of determining $View(\mathcal{V},\ell)$ merely from what "$\mathcal{X}$ knows initially". Finally, the function h shows exactly how "what $\mathcal{X}$ knows" can change with

[4] The analogy between the basis of a vector space and a complete set of covert channels isn't perfect, however. There is no notion of "dimension" as such for complete channel sets because the number of channels in a complete set can vary for a given multilevel system.

time. The fact that h is one-to-one indicates that the only thing that "$\mathcal{X}$ ever knows" is what "$\mathcal{X}$ knew initially" plus what "$\mathcal{X}$ learned" from fully and carefully observing the output "without ever missing a trick".

With the above idea of how to model a fully informed onlooker in our "closed world", we can formally define our Entropy Conservation Law:

Theorem III: The Entropy Conservation Theorem [Second Form of the Entropy Conservation Law]. Consider any (closed) multilevel system operating over some closed time interval $[t_1, t_2]$. Let $\mathcal{X}$ be (the random variable denoting) a fully informed onlooker with clearance ℓ and let $\mathcal{X}_1$ and $\mathcal{X}_2$ be abbreviations for $\mathcal{X}(\mathcal{V}, \ell, t)$ at the respective endpoints for $[t_1, t_2]$, and let Out_1 and Out_2 be similar abbreviations for the variable $Output(\mathcal{V}, \ell, t)$ at their respective times in $[t_1, t_2]$. Let $\Delta H_{out}(\mathcal{X})$ be the total change in system output entropy perceived by $\mathcal{X}$ over $[t_1, t_2]$. Let C represent any set of covert channels (with C_1 and C_2 corresponding to similar abbreviations for the ensemble covert channel $C(\mathcal{V}, \ell, t)$) and let $\Delta H_C(\mathcal{X})$ be that part of the output entropy reduction over $[t_1, t_2]$ due to action of the covert channels in C as perceived by $\mathcal{X}$. Let the term ξ_C denote what is called the *noise-effect entropy*, which is defined by $\xi_C = H(Out_2 | C_2, Out_1)$, and is the equivocation between the ending system output and the covert channels in C acting over $[t_1, t_2]$ (given Out_1 and also $View(\mathcal{V}, \ell)$; not shown). Then, the Entropy Conservation Law states that the covert channels in C are complete if and only if for *every* fully informed onlooker $\mathcal{X}$:

$$\Delta H_{out}(\mathcal{X}) \quad = \quad \Delta H_C(\mathcal{X}) \quad + \quad \xi_C$$

Proof: By definition we have:

$$\Delta H_{out}(\mathcal{X}) \quad = \quad H(Out_2 | \mathcal{X}_1) - H(Out_2 | \mathcal{X}_2)$$

but $H(Out_2 | \mathcal{X}_2) = 0$, since $\mathcal{X}$ fully learns the output at time t_2, thus:

$$\Delta H_{out}(\mathcal{X}) \quad = \quad H(Out_2 | \mathcal{X}_1)$$

Adding, and then subtracting $H(Out_2 | C_2, \mathcal{X}_1)$ to the right side and regrouping gives:

$$\begin{aligned} \Delta H_{out}(\mathcal{X}) \quad = \quad & [\, H(Out_2 | \mathcal{X}_1) \; - \\ & \qquad H(Out_2 | C_2, \mathcal{X}_1) \,] \\ + & \quad H(Out_2 | C_2, \mathcal{X}_1) \end{aligned}$$

But the expression in [] is just the entropy change due to covert channels over $[t_1, t_2]$, so this gives:

$$\Delta H_{out}(\mathcal{X}) \quad = \quad \Delta H_C(\mathcal{X}) \quad + \quad H(Out_2 | C_2, \mathcal{X}_1)$$

By our model for $\mathcal{X}$, $\mathcal{X}_1$ has information content equivalent to the pair $< Out_1, \mathcal{X}_0 >$, where $\mathcal{X}_0$ is the prior information about $\mathcal{V}$ known to onlooker $\mathcal{X}$, so the preceding is equivalent to:

$$\Delta H_{out}(\mathcal{X}) \quad = \quad \Delta H_C(\mathcal{X}) \quad + \quad H(Out_2 | C_2, Out_1, \mathcal{X}_0)$$

Finally, by the definition of completeness for covert channels, the complex entropy term to the right of "+" above is equivalent to the noise-effect entropy ξ_C if and only if the covert channels in C are complete.
□

The preceding Theorem is regarded as an Entropy Conservation Law since it turns out that if a set of covert channels is *not* complete, then it will *always* be the case that we can find *some* "fully informed onlooker" $\mathcal{X}$ so that:

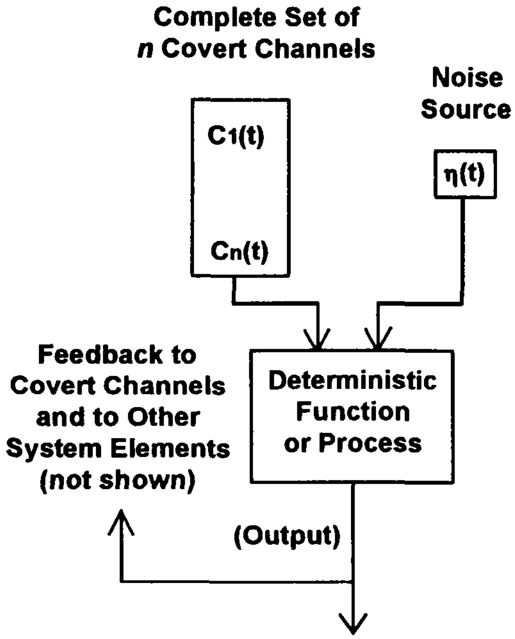
$$\Delta H_{out}(\mathcal{X}) \quad < \quad \Delta H_C(\mathcal{X}) \quad + \quad \xi_C$$

so that the output entropy is "not conserved" or "gets lost somewhere". The noise-effect entropy ξ_C represents the maximum amount by which $\Delta H_{out}(\mathcal{X})$ can differ from $\Delta H_C(\mathcal{X})$, and it is only possible to achieve the maximum difference ξ_C (for *all* fully informed onlookers) when the underlying set of covert channels is complete. However, with the current form of the Entropy Conservation Law, it is difficult to explain in English a simple physical interpretation for the noise-effect entropy term ξ_C, so I will postpone doing so until after presenting the idea of Covert Channel Normal Form.

4.2 Covert Channel Normal Form (CCNF)

The Entropy Conservation Law can also be expressed graphically with the following:

Informal Description of Covert Channel Normal Form [Third Form of the Entropy Conservation Law]. Consider some closed multilevel information system and any complete set of covert channels C for that system. It is always possible to express the output of the system (seen at a given security level) in *Covert Channel Normal Form* (CCNF) by using *only* the covert channels in C and a noise generator to drive some deterministic function or process (as shown by this next picture):



The previous picture, and the Covert Channel Normal Form concept generally, will now be formalized. Recall that I've defined covert channels so that their behavior is some property of (determined by) the system output. This formalization of the Entropy Conservation Law (Covert Channel Normal Form) is a sort of "weak converse" in saying that the system output is determined by any complete set of covert channels to within some noise level:

Theorem IV: The Covert Channel Normal Form Theorem [Fourth Form of the Entropy Conservation Law]. Consider any (closed) multilevel information system operating over a time interval $[t_1,t_2]$. Consider any set of covert channels $C_1,....,C_n$ with an ensemble covert channel $C(\mathcal{V},\ell,t)$. Then, the channels $C_1,....,C_n$ are complete *if and only if* there exists a deterministic function *f*, and a noise process $\eta(t)$ such that $(t_2 \geq t_1)$:

$$
\begin{aligned}
Output(\mathcal{V},\ell,t_2) \;=\; & \\
f\;(\; & Output(\mathcal{V},\ell,t_1), \\
& C(\mathcal{V},\ell,t_2), \\
& View(\mathcal{V},\ell), \\
& \eta(t_2),\; t_1,\; t_2 \;\;)
\end{aligned}
$$

Proof: First, we suppose that f exists and use the Second Form of the Entropy Conservation Law (the Entropy Conservation Theorem) to prove the channels are complete. Adopting a similar notation as in the proof of the Second Form, we consider:

$$
\begin{aligned}
\Delta H_{out}(\mathcal{X};t_1 \rightarrow t_2) \;=\; & \\
& H(Out_2|Out_1 = \lambda, \mathcal{X}_0 = \nu) \;\; - \\
& H(Out_2|Out_2 = \lambda', \mathcal{X}_0 = \nu)
\end{aligned}
$$

where $\Delta H_{out}(\mathcal{X};t_1 \rightarrow t_2)$ refers to some *particular* conditional entropy change over interval $[t_1,t_2]$ due to a specific "random output trial" (not the "average entropy change" as in the Second Form above). Note that we have used our "fully informed onlooker" model and expanded the terms $\mathcal{X}_1$ and $\mathcal{X}_2$ into the initial knowledge $\mathcal{X}_0$ and the corresponding output variable. Also, note that the second term is zero, so:

$$
\Delta H_{out}(\mathcal{X};t_1 \rightarrow t_2) \;=\; H(Out_2|Out_1 = \lambda, \mathcal{X}_0 = \nu)
$$

Now further consider a *specific* random action taken by our covert channels up to time t_2 written $C_2 = \alpha$. Then adding and subtracting the conditional entropy and re-grouping, we have:

$$
\begin{aligned}
\Delta H_{out}(\mathcal{X};t_1 \rightarrow t_2) \;=\; & \\
[\;\; & H(Out_2|Out_1 = \lambda, \mathcal{X}_0 = \nu) \;\; - \\
& \quad H(Out_2|C_2 = \alpha, Out_1 = \lambda, \mathcal{X}_0 = \nu)\;] \\
+ \;\; & H(Out_2|C_2 = \alpha, Out_1 = \lambda, \mathcal{X}_0 = \nu)
\end{aligned}
$$

Now by assumption, the function *f* determines output at t_2 given C_2 and Out_1 independently from $\mathcal{V}$, and therefore, also independently of $\mathcal{X}_0$. Thus, in our particular random trial, we can eliminate $\mathcal{X}_0$ in the term to the right of the "+" giving:

$$
\begin{aligned}
\Delta H_{out}(\mathcal{X};t_1 \rightarrow t_2) \;=\; & \\
[\;\; & H(Out_2|Out_1 = \lambda, \mathcal{X}_0 = \nu) \;\; - \\
& \quad H(Out_2|C_2 = \alpha, Out_1 = \lambda, \mathcal{X}_0 = \nu)\;] \\
+ \;\; & H(Out_2|C_2 = \alpha, Out_1 = \lambda)
\end{aligned}
$$

Now we can multiply both sides of the above by the joint probability $pr(C_2 = \alpha, Out_1 = \lambda, \mathcal{X}_0 = \nu)$ and sum over all values (all random trials) giving:

$$
\begin{aligned}
\Delta H_{out}(\mathcal{X}) \;=\; [\;\; & H(Out_2|Out_1, \mathcal{X}_0) \;\; - \\
& \quad H(Out_2|C_2, Out_1, \mathcal{X}_0)\;] \\
+ \;\; & H(Out_2|C_2, Out_1)
\end{aligned}
$$

But this is the same as:

$$
\Delta H_{out}(\mathcal{X}) \;=\; \Delta H_C(\mathcal{X}) \;+\; \xi_C
$$

which, from the Entropy Conservation Theorem (Theorem III) we can conclude that the covert channels $C_1,....,C_n$ are complete. Conversely, suppose the covert channels $C_1,....,C_n$ are complete. We define (an algorithm for) a deterministic function *f* as follows:

```
algorithm f:
  input values for $t_1$ and $t_2$
  input value for $Output(\mathcal{V},\ell,t_1)$
  input value for $C(\mathcal{V},\ell,t_2)$
  input value for $View(\mathcal{V},\ell)$
  input (random) value $0 < \eta(t_2) < 1$
  compute
    the conditional probability distribution for
    variable $Output(\mathcal{V},\ell,t_2)$ given
        $Output(\mathcal{V},\ell,t_1)$;
        $C(\mathcal{V},\ell,t_2)$;
        $View(\mathcal{V},\ell)$
    and call this measure $pr^*$. Now, assuming that the
    values of variable $Output(\mathcal{V},\ell,t_2)$ are well-ordered
    in some way, we next........
  compute
    the least value of $x$ such that the cumulative
    distribution function for $pr^*$ equals or exceeds the
    random input value from the noise function:
```

$$\eta(t_2) \quad \leq \quad \sum_{\tau \leq x} pr^*(Output(\mathcal{V},\ell,t_2) = \tau)$$

```
  output $x$ as the value of the function f.
```

Because by assumption, the covert channels associated with $C(\mathcal{V},\ell,t)$ are complete, the distribution for $Output(\mathcal{V},\ell,t_2)$ will be independent of variable $\mathcal{V}$ given $C(\mathcal{V},\ell,t_2)$, so the above function f will produce the proper joint distribution for $Output(\mathcal{V},\ell,t_2)$ and variable $\mathcal{V}$. Thus the function f just described is a function having the required properties.
□

The uncertainty in the output of deterministic function f in the Covert Channel Normal Form (CCNF) obviously comes from two sources: (1) uncertainty from the set of covert channels and (2) uncertainty from the noise process $\eta(t)$. These two separate sources of output uncertainty show the connection between CCNF and the more abstract (First and Second) forms of the Entropy Conservation Law, and also explains why I refer to ξ_C as the noise-effect entropy. Basically, the noise-effect entropy ξ_C in the First and Second forms of the Entropy Conservation Law is sort of the "image of noise function $\eta(t)$ under function f" for a system's representation in CCNF. Of course, this relationship between $\eta(t)$ and ξ_C is only sensible precisely when the underlying covert channel set is complete. The comparison is meaningless otherwise, because the "image of $\eta(t)$ under f" given the output of an incomplete set of covert channels would be correlated with sensitive information in $\mathcal{V}$, and thus could not be related to the noise-effect entropy ξ_C.

5 Thoughts on Potential Applications

The Entropy Conservation Law indicates that the question about completeness of a set of covert channels comes down to whether or not the system output can be represented in CCNF using only the known covert channels. This suggests that given an arbitrary multilevel system, one might attempt to search for a complete set of covert channels with the following "algorithm":

```
Engineering Paradigm I:

  Channel_Set ← ∅;
  N ← 0;

  while
     the system output (at a given level) cannot be represented in
     Covert Channel Normal Form using only the (known) covert
     channels contained in "Channel_Set"
  do
        N ← N + 1;
        Channel_Set ← Channel_Set ∪ {$C_N$};
                [where $C_N$ is some covert channel not already in
                "Channel_Set"]
  end loop

  output   Channel_Set; [Complete set]
```

Here again we have an analogy with vector spaces. The above is very similar to extending a set of vectors to a basis [Halmos74], except we haven't defined any notion of "independence of covert channels" as such. The Entropy Conservation Law guarantees that if the above "algorithm" terminates, then it will determine a correct (complete) set of covert channels in the formal sense. However, use of the above "algorithm" for covert channel analysis must deal with several (related) problems, for any given system under analysis:

1. There may not be an easy way of deciding whether or not a CCNF exists for a given set of channels.

2. Even if a CCNF decision procedure exists, there may be no easy way of locating a new covert channel when the decision procedure rejects a candidate set of covert channels.

3. The "algorithm" in theory may not terminate. We would like to be sure that every incomplete set can be finitely extended to a complete set.

These three problems seem serious enough so that finding a CCNF for an *arbitrary* multilevel system is problematic. It makes sense therefore to focus on *particular* security architectures for which a CCNF is readily apparent; this suggests a departure from the above engineering paradigm.

5.1 Designing Systems to Simplify Channel Analysis

In as much as one might design a system to be reliable, maintainable, testable, or to support formal proof, one might also design a system with the specific goal in mind of simplifying covert channel analysis. So rather than trying to find the CCNF for an *arbitrary* multilevel secure system, I suggest that one could limit the engineering paradigm to pursue *specialized* security architectures for which a useful CCNF is easily found:

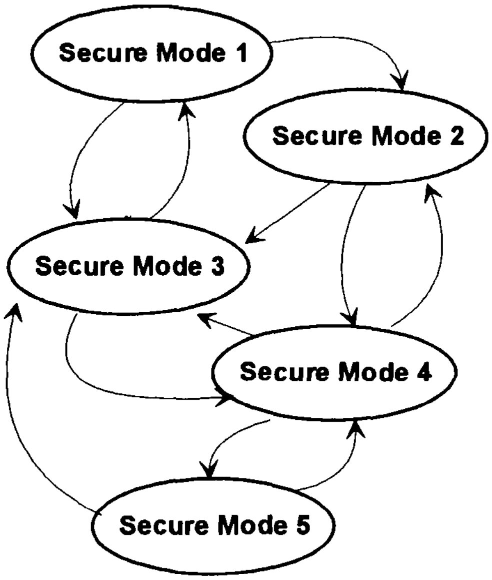
Engineering Paradigm II:

```
define
    a security architecture for a specific security application
prove
    a "normal form theorem" for the particular security architecture
    so that any system that satisfies the architecture will always be
    guaranteed to have a well-understood CCNF
build
    a system that satisfies the chosen security architecture
verify
    that the system satisfies the chosen security architecture
analyze
    the covert channels in the system relying mainly on the "normal
    form theorem" for the architecture
```

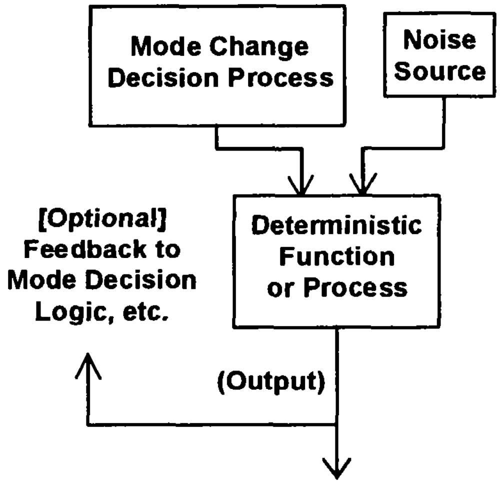
The above engineering paradigm is thus a change from doing covert channel analysis based mainly on system peculiarities, to more emphasis on covert channel analysis at a higher level of architecture using a more general understanding of the covert channel characteristics of a given security architecture. I now illustrate how such an engineering paradigm might be used by discussing a specific security architecture that has a particularly simple CCNF.

5.2 Mode Security

In [Browne94], a security architecture known as Mode Security is presented[5]. The idea behind Mode Security is to partition a multilevel system's states into disjoint sets called *secure modes*. This is done in a way so that the apparent behavior of a Mode Secure system is that the system will spend most of its time fixed in one of it modes. Periodically, the machine may "jump" from a secure mode to another, and [Browne94] represents this "periodic jumping" by an automaton showing how the secure modes can change. This next picture shows an example of such an automaton (called a Characteristic Automaton in [Browne94]) having five secure modes:



The basic claim in [Browne94] is that this organization of the system into secure modes ensures that all covert channels are connected with mode change events. Supposedly, there are never any covert channels between any two consecutive mode change events. We can support this claim by the fact that every Mode Secure system has the following general CCNF:

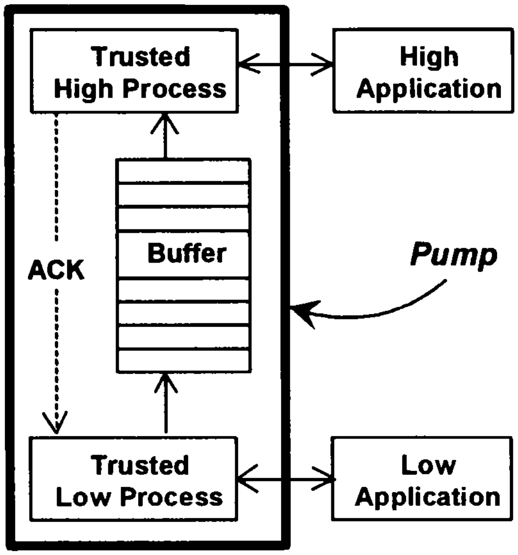

Mode Change Decision Process
Noise Source
[Optional] Feedback to Mode Decision Logic, etc.
Deterministic Function or Process
(Output)

In the above CCNF, the noise process is the only thing driving the (otherwise deterministic) system output while the system operates in one of its secure modes. The mode change decision process is idle and produces no output unless and until a mode change occurs. The fact that all Mode Secure systems have the above CCNF can be proven by Theorem B of [Browne94], together with Theorem I and the Covert Channel Normal Form Theorem (Theorem IV) in this paper (although I will not give the proof, here). This supports the claim in [Browne94] that all covert channels within a Mode Secure machine are associated with mode change decisions, and there are no other kinds of covert information flow in such a system.

[5] I emphasize that Mode Security is *not* a covert channel analysis method; it is a *specialized architecture* (having covert channels that may be analyzed a variety of ways). Mode Security is *not* a panacea; it is *only* intended for certain security applications [Browne94], so the usefulness of finding a CCNF for a Mode Secure system is limited to those applications for which Mode Security itself is useful. This is exactly the sort of specialization of the application domain that I've suggested that is probably needed to facilitate a search for a CCNF.

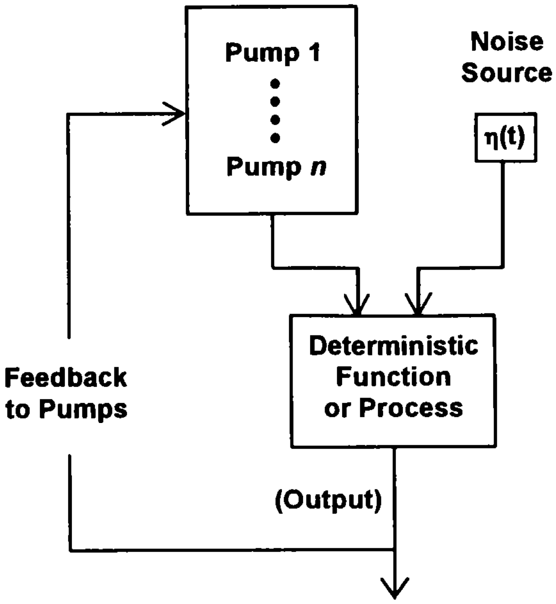
5.3 The Kang-Moskowitz Data Pump

To further illustrate specialized uses of CCNF, one might like to devise a security architecture that makes deliberate allowances for covert channels to serve some specific purpose. Covert channels might be deliberately allowed in a system to provide reliable multilevel communication, for instance. One way to do this would be to use a Pump [KangMoskowitz93] for efficient, reliable, and (adequately) secure communication between a pair of applications at two different security levels:



Each Pump consists of a pair of trusted "low" and "high" processes that share a buffer. The Pump permits reliable inter-level communication by permitting "high" to "low" acknowledgments under the auspices of the Pump's internal trusted processes. The internal acknowledgment from the trusted "high" process to the trusted "low" process (dashed, downward arrow in the above picture) is implicit; the actual mechanization for the acknowledgment [KangMoskowitz93] is having the trusted "low" process detect when space is available in the buffer. The Pump controls covert timing channels by having the trusted "low" process inject random time delays into its own acknowledgments to the "low" application external to the Pump.

One might implement a trusted operating system which provides a set of n of these Pumps for use by application processes. Now suppose that we wanted to prove that the *only* covert channels in our system were due to the action of these Pumps. Then, from our Entropy Conservation Law, we could prove our claim if we could somehow show that our architecture had the following CCNF:



The above picture is just for illustration since, without looking at a specific architecture, one cannot know what design assumptions would be needed to ensure that the n Pumps were the sole source of covert channels. What we *do* know is that if the architecture could not be put into the above CCNF, we could be certain from the Entropy Conservation Law that the set of n Pumps did not constitute a complete set of covert channels (and this would reveal a hidden covert channel). We could then identify and document the missing covert channel(s), or redesign our system to remove those covert channels not related to action of the Pumps and again attempt to put our system into CCNF to verify our redesign effort.

6 Remarks

In the preceding section, we discussed issues pertaining to applications of the Entropy Conservation Law. In this section, I make other observations about the Entropy Conservation Law as well as discussing new topics mainly relating to open problems.

6.1 Interpreting the "Toy Problem"

We return to our 2-input/3-output "toy" problem of Section 2. If we form a closed system (by tying the two inputs to a pair of SECRET input generators and embed those generators as part of our system), the Entropy Conservation Law requires that any complete set of covert channels must constrain all three outputs. We can put our "toy" problem in CCNF for any set of channels that determine the value of all three outputs. (In this case, the noise-effect entropy ξ_C is zero).

6.2 More about the Scope Problem

The fact that we have managed to separate the scope problem (identifying complete channel sets) from the quantification problem (computing the global covert information flow) is a "double-edged sword". The "toy problem" of Section 2 illustrates the bothersome situation that one can have a *complete* set of covert channels *all* of which have *zero channel capacity*, and still have an *overall* covert information flow with *non-zero* capacity. It isn't that the Entropy Conservation Law is wrong; the problem is that the notion may not always be helpful. In practice, the separation between the scope and quantification problems may not be as pure as we have achieved in theory.

In some cases for example, it may be that the only way to prove that a CCNF does not exist is to actually find a channel of non-zero capacity. In the case of Mode Security (described briefly in Section 5.2), it was very easy to separate the quantification and scope problems; specifically, Theorem B in [Browne94] gives a constructive proof that (in effect) the random process that drives mode changes is, by itself, a complete set of covert channels, and the proof in no way depended on actually calculating the global covert information flow (the latter which is done by Theorem C in [Browne94]). However, it is not clear if this is entirely an accidental aspect of Mode Security, or if there really is a natural separation, in general, between proving channel completeness and quantifying covert information flow. Given all of this, more work is clearly needed to see if separating the scope and quantification problems would be useful in practice.

6.3 Comments about Noise-Effect Entropy

Theorems I-IV show that completeness of a covert channel set is equivalent to whether the set satisfies the Entropy Conservation Law in its various forms. If the Entropy Conservation Law is *not* satisfied for a given set of channels, there is no question that *some* covert channel *is missing* from the set. On the surface, we are thus immune from the effect of "formal flows" [Eckmann94] in the sense that violations of the Entropy Conservation Law *always* indicate *some genuine* compromising covert information flow that has been omitted from a given covert channel set. Using the Entropy Conservation Law, we are ostensibly never unsure whether some unexplained behavior of the system output is due to a real covert channel or a "phantom" effect due to an overly conservative security policy; unexplained system output behavior is *always* a covert channel, if not noise.

Nevertheless, there is a problem with the Entropy Conservation Law that is analogous to the problem of "formal flows". The unfortunate fact is that the noise-effect entropy ξ_C is not unique. As the subscript C in ξ_C indicates, the "residual output noise" does depend on the set of covert channels C and can differ between complete sets. Informally, complete sets of channels with larger values of ξ_C are "better" than those with lower values, since, the more that output entropy can be attributed to noise, the less that output entropy need be attributed to covert channels. The current theory does not address this at all, however.

6.4 Quasi-Complete Covert Channel Sets

In lieu of finding complete sets of covert channels, one might wish to develop a theory whereby one could compute the "residual global flow" for an incomplete set of channels as a "goodness" measure for covert channel analysis. Our Entropy Conservation Law sheds some light on this since, for a given set of covert channels $C_1, \ldots . C_n$, when the expression:

$$\max_{\mathcal{X}} \left(\xi_C - H(output|C_1, \ldots ., C_n, \mathcal{X})\right)$$

is small, then the covert channels $C_1, \ldots . C_n$ are in a sense "close" to being a complete set. One might prove, for a particular security architecture, some kind of "Remainder Theorem" that gives an upper bound on that part of the global information flow not captured by some nearly complete covert channel set $C_1, \ldots . C_n$.

6.5 Informal Measures of Completeness

Yet another idea that our work has cast some light on, would be a result that shows that if one identifies at random a large number of zero or low capacity covert channels, then one might put a rigorous upper bound on the probability of finding a channel having a capacity exceeding some value ε. The reason I suspect this is not epistemological, but purely mathematical. Consider the following result:

> **Theorem V.** For every pair of natural numbers $n, m > 0$, there exists a system with n input channels and m output channels such that: (1) *all* channels with *fewer* than m outputs (if any) have zero capacity, and (2) there exists at least one channel of non-zero capacity (having m outputs).

This theorem, which amounts to a generalization of our "toy" problem, has a simple constructive argument whereby one constructs 2^n tables with 2^m rows each (with $n = 2$ and $m = 3$ in Section 2) for giving the conditional probability of each of 2^m output combinations for every 2^n possible inputs. Initially, every one of the 2^m rows in all 2^n tables is set to a uniform distribution $1/2^m$ making all channels have zero capacity. To get a channel of non-zero capacity, simply add the amount $((k-1)/(2^{m+1} * 2^n))$ to every table row in table number k when the row corresponds to an output combination having an even number of "1" outputs; and subtract that same amount from every row corresponding to an odd number of "1" outputs. The point to this construction is that, the channel capacity of the "global channel" that is constructed is quite small and seems to shrink as $n, m \mapsto \infty$. This appears to be generally true; mathematically, if one wants to "hide" a statistical dependency among a large set of random variables, such that all medium and smaller size sets are zero or low capacity channels, it seems that the amount of probability mass that one can "redistribute" (via a perturbation argument such as I just gave) tends to be limited and this seems to have a substantial mitigating effect on potential global channel capacity. There thus seems to be a mathematical basis for concluding that global covert information flow is small when large numbers of complex, small capacity channels are found via extensive search.

7 Conclusion

In this paper, we used a simple formal theory of covert channels to formalize the notion of a *complete* set of covert channels. The purpose in this was to define a formal test for determining when enough individual covert channels have been discovered so as to capture global covert information flow.

Complete sets of covert channels turn out to be precisely those that satisfy a certain Entropy Conservation Law. This Entropy Conservation Law was expressed in different forms; most notable is the notion of a Covert Channel Normal Form (CCNF) for a multilevel system. The existence of a CCNF, specifically, as well as the Entropy Conservation Law, generally, were shown to entail a necessary and sufficient test for whether or not a further search for covert channels is warranted.

Finally, we touched on possible connections between theory and practice while recognizing that there are numerous open issues. While the Entropy Conservation Law is theoretically "universal", as a practical matter, the use of the Entropy Conservation Law would probably have to be limited as described in Section 5. I suggested that rather than trying to apply the Entropy Conservation Law directly to an *arbitrary* covert channel analysis problem, it would probably be better to apply the Entropy Conservation Law only to *specific* covert channel analysis problems. I indicated that specialized use of the Entropy Conservation Law could mean designing specific security architectures for which the concepts of entropy conservation (such as CCNF) could be directly applied to such systems, and some suggestions were made along these lines. A number of open areas relating to the results in the paper were also identified.

8 Acknowledgments

I would like to thank Dr. R. E. Newman-Wolfe of the University of Florida, and Dr. Ira Moskowitz of the U.S. Naval Research Laboratory for various interesting discussions that led to, and illuminated many of the ideas presented in this paper. Dr. Moskowitz' comments on drafts of this paper were very helpful, along with those of the referees. Also, I would like to thank Dr. Jim Gray of the Hong-Kong University of Science and Technology, and also Mr. Stanley Perlo, for numerous useful discussions over a period of several years that have had a strong influence on this author's thinking about security issues, and for providing many useful insights into the theory of covert information flow.

References

[Browne91] Browne, R., "The Turing Test and Non-Information Flow", *In Proceedings of the 1991 IEEE Symposium on Research in Security and Privacy.*

[Browne94] Browne, R., "Mode Security: An Infrastructure for Covert Channel Suppression", *In Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy.*

[Eckmann94] Eckmann, S., "Eliminating Formal Flows in Automated Information Flow Analysis", *In Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy.*

[GoguenMeseguer84] Goguen, J., and Meseguer, J., "Unwinding and Inference Control", *In Proceedings of the 1984 IEEE Symposium on Security and Privacy.*

[Halmos74] Halmos, P., "Finite-Dimensional Vector Spaces", *Springer-Verlag, 1974.*

[KangMoskowitz93] Kang, M., and Moskowitz, I., "A Pump for Rapid, Reliable, Secure Communication", *In Proceedings of the 1st ACM Conference on Computer and Communications Security, November, 1993.*

[NCSC93] "A Guide to Understanding Covert Channel Analysis of Trusted Systems", *Security Guideline issued by the National Computer Security Center, Document Number NCSC-TG-030, Version 1, November 1993.*

[Rushby81] Rushby, J., "The Design and Verification of Secure Systems", *In Proceedings of the 8th ACM Symposium on Operating System Principles, December, 1981.*

[Shannon48] Shannon, C., "A Mathematical Theory of Communication", *Bell System Technical Journal, Volume 27, July 1948.*

[Shannon61] Shannon, C., "Two-Way Communication Channels", *In Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics, and Probability, Volume 1, 1961.*

[WittboldJohnson90] Wittbold, J.T., and Johnson, D., "Information Flow in Non-deterministic Systems", *In Proceedings of the 1990 IEEE Symposium on Research in Security and Privacy.*