APN Functions and all that jazz

Gary McGuire

School of Mathematical Sciences
University College Dublin
Ireland
and
Claude Shannon Institute
www.shannoninstitute.ie

SORIA July 2010

Sketch of talks:

1. Talk 1: S-Boxes, Differential Attack, APN Functions, Coding Theory, Equivalence
2. Talk 3: Linear Attack, Fourier Spectrum, Nonlinearity
3. Talk 4: Other topics, Supersingular Curves, Other Groups, Stickelberger's Theorem

Talk 1

1. APN definition
2. Permutations, Quadratic Functions
3. S-Boxes, Differential Attack
4. Coding Theory
5. Equivalence

Definitions

Definition (Perfect Nonlinear function)

Let A, B be finite abelian groups, written additively, of the same cardinality. We say $f : A \to B$ is a perfect nonlinear (PN) function iff $f(x+a) - f(x) = b$ has at most one solution for all $a \in A$, $a \neq 0$, and all $b \in B$.

Equivalently:

- $f(x+a) - f(x) = b$ has exactly one solution,
- $f(x+a) - f(x)$ is bijective, for all nonzero a, (derivatives)
- The set $\{f(x+a) - f(x) : x \in A\}$ has cardinality $|A|$,
- $f(x+a) - f(x) = f(y+a) - f(y) \implies a = 0 \text{ or } x = y$.
- $f(z) - f(x) - f(w) + f(y) = 0$ and $z - x - w + y = 0$ $\implies x = z, y = w \text{ or } x = y, z = w$.

PN functions are also called planar functions if $A = B = \mathbb{F}_q$.

Example: $f(x) = x^2$ on a finite field of odd characteristic.

Definitions

PN functions do not exist in characteristic 2, because if x is a solution to $f(x+a)-f(x)=b$ then so is $x+a$. ☹
This is why the following definition is made.

Definition (Almost Perfect Nonlinear function)

Let A, B be finite abelian groups, written additively, of the same cardinality. We say $f : A \to B$ is an almost perfect nonlinear (APN) function iff $f(x+a)-f(x)=b$ has at most two solutions for all $a \in A$, $a \neq 0$ and all $b \in B$.

Mostly of interest when $A = B =$ finite field of order 2^n.
But note, the definition is additive in nature.
In char 2, the term "almost" is not justified.

Definitions due to Meier-Staffelback, Nyberg, early 1990s.

Definitions

Definition (Almost Perfect Nonlinear function)

Let A, B be finite abelian groups, written additively, of the same cardinality. We say $f : A \to B$ is an almost perfect nonlinear (APN) function iff $f(x+a) - f(x) = b$ has at most two solutions for all $a \in A$, $a \neq 0$ and all $b \in B$.

Mostly of interest when $A = B =$ finite field of order 2^n.
Equivalently, in that case,

- $f(x+a) - f(x)$ is 2-1, for all nonzero a,
- The set $\{f(x+a) + f(x) : x \in GF(2^n)\}$ has cardinality 2^{n-1},
- $f(x+a)+f(x) = f(y+a)+f(y) \implies a = 0 \text{ or } x = y \text{ or } x = y+a$
- $f(z) + f(x) + f(w) + f(y) = 0$ and $z + x + w + y = 0$
$\implies x = z, y = w \text{ or } x = y, z = w \text{ or } x = w, y = z$

Example: $f(x) = x^3$ on any finite field.
What about x^5 on $GF(2^n)$? EXERCISE

Linearized Polynomials

Recall linearized polynomials in $\mathbb{F}_q[x]$ have the form

$$L(x) = \sum a_i x^{q^i}$$

and their roots form a vector space U over $\mathbb{F}_q$. (you check this)
Suppose $U \cap \mathbb{F}_{q^n}$ has dimension k (over $\mathbb{F}_q$).
Then the equation $L(x) = c$ has either 0 or q^k solutions. (why?)

Bijective Functions

Suppose you wanted one of these functions to be bijective.
(This has certain applications, see later)

Theorem

PN bijections do not exist.

Proof: Let f be a PN function. Choosing b to be 0, for all nonzero a there must exist a solution to $f(x+a)-f(x)=0$.
Therefore, f cannot be bijective. $\square$

What about APN bijections? Do they exist?

APN Permutations

It depends on the group. On some groups, APN permutations do exist. (More on this later.)

Open Problem: Do APN permutations exist on finite fields $GF(2^n)$ where n is even?
(Remember x^3 is bijective iff n is odd)

$n = 4$ was checked by exhaustive computer search - none found.

On July 14, 2009, at the Fq 9 conference in Dublin, John Dillon announced an APN permutation on $GF(64)$!!

Quadratic, Bilinear Forms

Let $L = GF(p^n)$.
A function $f : L \longrightarrow L$ is called *quadratic* if, when f is written as a univariate polynomial, all exponents have the form $p^i + p^j$. In other words

$$f(x) = \sum_{i,j} a_{ij} x^{p^i + p^j}.$$

These are also called Dembowski-Ostrom polynomials.
Equivalently, $f(x+y) - f(x) - f(y)$ is linearized in both x and y (viz. $(x+y)^{p^i+p^j} = x^{p^i+p^j} + x^{p^i} y^{p^j} + x^{p^j} y^{p^i} + y^{p^i+p^j}$).

If $p = 2$, $\mathrm{tr}(f(x))$, or indeed $\mathrm{tr}(bf(x))$, is an $\mathbb{F}_2$-valued quadratic form.
Such a quadratic form is called *balanced* if it has $|L|/2$ zeros.
If not balanced, the number of zeros must be $2^{n-1} \pm 2^{(n-2+w)/2}$.

Remark

Sometimes we view L as an n-dimensional vector space over $\mathbb{F}_2$. Fixing a basis, one can write f as a polynomial in n Boolean variables:

$$f(x_1, \ldots, x_n) = \sum a_J x^J$$

where $x^J = x_1^{j_1} \cdots x_n^{j_n}$, $a_J \in \mathbb{F}_2$.

The algebraic degree of f is the usual degree of this polynomial. (called Algebraic Normal Form)

Start of a Proof of APN property

To establish the APN property of a function f on L we must show that the equation

$$f(x) + f(x+q) = p$$

has at most two solutions in L for all $p \in L$ and all $q \in L^*$.
If f is quadratic, then the number of solutions is 0 or has the same size as the kernel of the $\mathbb{F}_2$-linear map

$$\Delta_{f,q}(x) := f(x) + f(x+q) + f(q).$$

One shows, somehow, that the kernel of $\Delta_{f,q}$ has size 2 for each non-zero q in the underlying field.
This is why quadratic functions are "easier" (but not easy).

Cryptographic Motivation

Symmetric Cryptosystems have the same key for encryption and decryption.

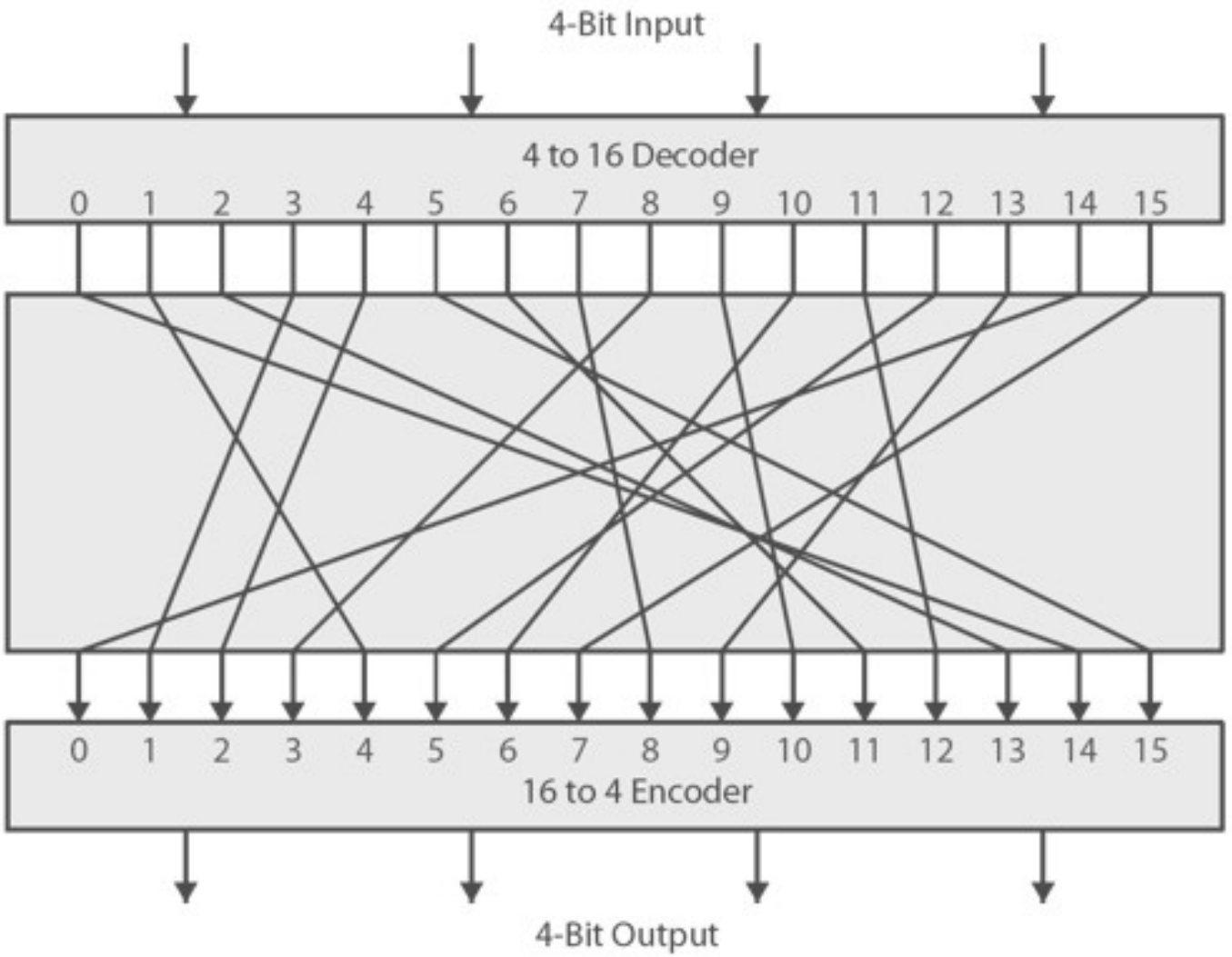
Claude Shannon introduced some design criteria.

He proposed "confusion and diffusion" in the encryption algorithm.

Shannon and Feistel introduced substitution-permutation symmetric cryptosystems to meet this requirement.

Many symmetric block ciphers now have a nonlinear part, called an S-Box. This provides the confusion.
A permutation provides the diffusion.

A 4-bit to 4-bit Substitution-Box, or S-Box.



This box represents a function from $\{0,1\}^4$ to itself.

This function should be nonlinear, as it provides the confusion.

Of course, this alone is easy to break.

How do we use this?
We take larger inputs and place several smaller S-boxes side-by-side.

A subkey is mixed in.

This is repeated several times, i.e., several rounds.
(One round is substitution + permutation + subkey)

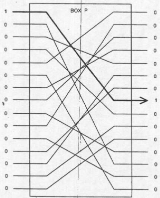
computer environment. Here the data transmitted may often be nonredundant if, for example, they are purely numerical, and an error in a single digit can cause an avalanche of computational errors. Study of the problem has shown that simple error-detecting codes are inadequate for guarding the integrity of computer data against possible tampering by a human expert. What is required is not mere error-detection but cryptographically protected authentication. Surprisingly, this is best achieved by relying on certain principles inherent in the cipher structure itself. Rather than trying to modify the stream concept, let us take a fresh look at the basis of all cryptography: substitution on blocks of message digits.

We shall refer to any cipher that converts n message digits into n cipher digits as a block cipher. For example, a block cipher would be one that turns 00000, standing for a clear-text A, into, say, 11001, the cipher equivalent of A according to some permutation key, exactly as a tableau does. To see how such a binary transformation is performed by an electronic device let us consider a substitution on only three binary digits [*see illustration on preceding page*].

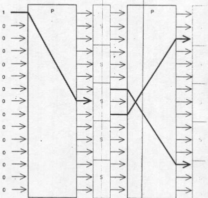
Three binary digits can represent eight items: 2^3 equals eight. The substitution device consists of two switches. The first converts a sequence of three binary digits into its corresponding value to the base eight, thereby energizing any one of eight output lines. These eight lines can be connected to the second switch in any one of 8!, or 40,320, ways. We are at liberty to decide which one of these 40,320 distinct connection patterns, or wire permutations, is to be made between the first switch and the second switch. The role of the second switch is to convert the input, presented as one digit to the base eight, back into a three-digit binary output.

If the substitution device were built to handle a five-digit binary input, it could be used to encipher an alphabet of 32 letters. The number of possible connection patterns between the two switches would then be 32!. That would seem to be an incredibly large number of keys, but the cipher produced must still be regarded as glaringly weak: it could not resist letter-frequency analysis. The weakness is not intrinsic; the device described is mathematically the most general possible. It includes, for any given input-output dimension, any possible reversible cipher that has been or ever could be invented; mathematicians would say it represents the full symmetric group. It is completely "nonsystematic": one permutation connection tells an opponent nothing at all about any other connection. The problem is not intrinsic, then, but is related to size. In spite of the large number of keys, the "catalogue" of possible inputs and outputs is too small: only 32. What is required is a catalogue so large that it is impractical for any opponent to record it. If we had a box with 128 inputs and outputs, for example, an analyst would have to cope with 2^{128} (or more than 10^{38}) possible digit blocks, a number so vast that frequency analysis would no longer be feasible. Unfortunately a substitution device with 128 inputs would also require 2^{128} internal terminals between the first and the second switch, a technological impossibility. This is a fundamental dilemma in cryptography. We know what would be ideal but we cannot achieve the ideal in practice.

Perhaps one could find a device that is easy to realize for a large number of inputs. One might, for example, build a box with, say, 128 input and 128 output terminals that are connected internally by ordinary wire crossings [*see illustration at left below*]. Such a "permutation box" with $n!$ terminals would have $n!$ possible wire crossings, each of which could be set by a different key. It could be built easily for $n = 128$. Although this provides a usefully large number of keys (128!), we are now faced with a new difficulty. By the use of special trick messages it is possible to read out the complete key to such a system in only $n - 1$ (in this case 127) trials. The trick



PERMUTATION BOX can handle very many terminals but it only shuffles positions of digits. An opponent can learn its wiring by feeding in inputs with single 1's and seeing where 1's come out.



PRODUCT-CIPHER SYSTEM combines P boxes and S boxes. The P boxes have a large number of inputs (represented by 15 in the illustration) and the S boxes a number that is manageable for such

is to introduce a series of messages containing a single 1 at $n - 1$ positions; the position of the 1 in the output betrays the particular wire crossing used in the box. The flaw in the simple permutation box is again that it is a linear system.

We need a compromise that will at least approximate the features of the general system. We are led to the notion of a product cipher in which two or more ciphers are combined in such a way that the resulting system is stronger than either of the component systems alone. Even before World War I various cumbersome ciphers using several stages of encipherment were studied. The first genuinely successful example was probably the one devised by the Germans that was known as the ADFGVX system. We need only observe here that it coupled "fractionation" with "transposition." By that procedure a message was broken into segments and the segments were transposed. The important fact to note here is that the result of a product cipher is again a block cipher; the goal, of course, is that the cipher behave as much as possible as if it were a general substitution cipher.

Between World War I and World War II interest in product ciphers almost totally disappeared because of the successful development of rotor, or wired-wheel, machines, which belong to the general class of pseudorandom-stream generators. A typical rotor machine has a keyboard resembling that of a typewriter. Each letter is enciphered by the operation of several wheels in succession, the wheels being given a new alignment for each new letter according to an irregular and keyed stepping algorithm. The message is decoded by an identical machine with an identical key setting.

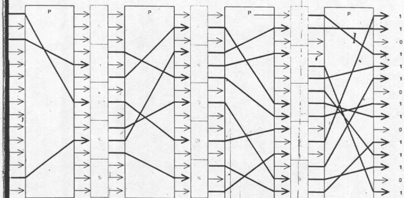
The modern interest in product systems was stimulated by a paper by Claude E. Shannon titled "Communication Theory of Secrecy Systems," published in the *Bell System Technical Journal* in 1949. In a section on practical cipher design Shannon introduced the notion of "mixing transformation," which involved a special way of using products of transformations. In addition to outlining intuitive guides that he believed would lead to strong ciphers, he introduced the concepts of "confusion" and "diffusion." The paper opened up almost unlimited possibilities to invention, design and research.

The manner in which the principles of confusion and diffusion interact to provide cryptographic strength can be described as follows. We have seen that general substitution cannot be realized for large values of n, say $n = 128$, and so we must settle for a substitution scheme of practical size. In the IBM system named Lucifer we have chosen $n = 4$ for the substitution box. Ev
though 4 may seem to be a small nu
ber, it can be quite effective if the su
stitution key, or wire-crossing pattern,
properly chosen. In Lucifer nonline
substitution effectively provides the e
ment of confusion.

We have also seen that a linear p
mutation box is easy to build even f
$n = 128$. The number of input and o
put terminals is simply equal to n. Bei
a pure digit-shuffler, a device that me
ly moves digits around without alteri
the number of 1's in the data, the p
mutation box is a natural spreader
confusion, that is, it can provide optim
diffusion.

In the Lucifer system the input da
pass through alternating layers of bo
es that we can label P and S. P stan
for permutation boxes in which n is
large number (64 or 128) and S stan
for substitution boxes in which n is sm
(4). Whereas either P boxes alone or
boxes alone would make a weak syste
their strength in combination is cons
erable.

One measure of strength is depicted
a device in which for simplicity the
boxes have $n = 15$ and the S boxes ha
$n = 3$ [*see illustration on these t*
pages]. If we imagine this sandwich
boxes being "tickled" by addressing
with a specially selected input, whi
might consist of a number made up

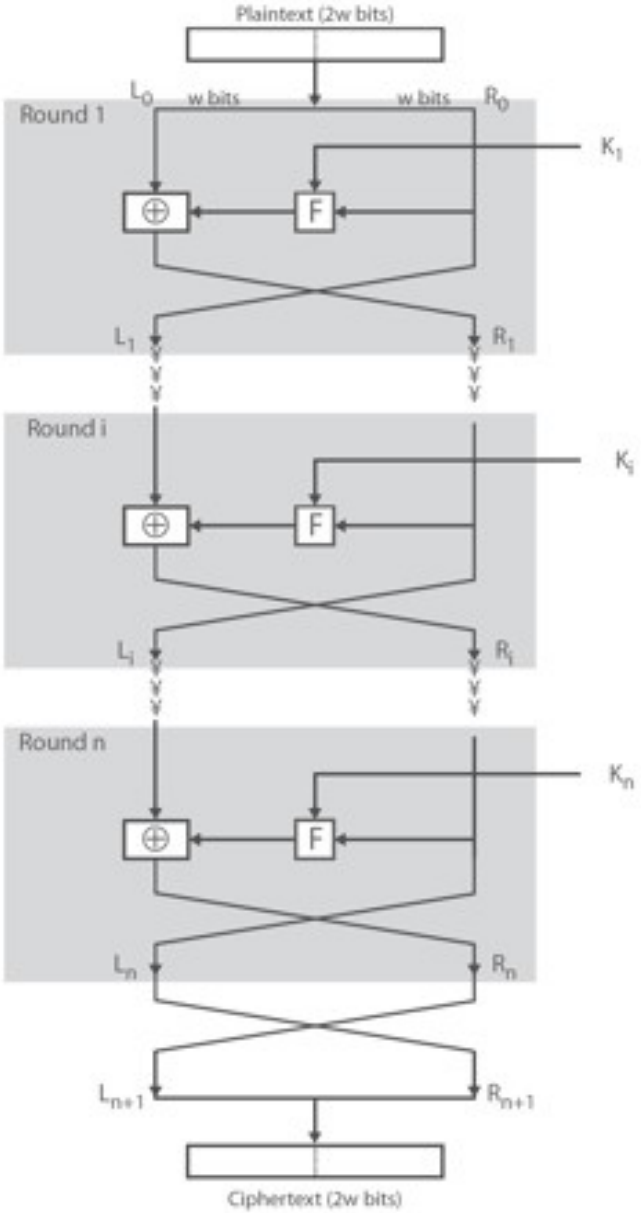


devices—three in this case. The P boxes shuffle the digits, providing "diffusion." The S boxes provide nonlinear substitution and thus "confusion." In this simplified example the input includes a single 1 and 14 0's. Because the S boxes are nonlinear, they can p
tentially increase the number of 1's; meanwhile the P boxes mo
the 1's around. The result can be an unpredictable avalanche of 1

DES (Data Encryption Standard, 1970s) has 48 bit input.
DES uses eight S-boxes which each take 6 bits for input, and have 4-bit output.
(The S-boxes are different)

DES is a Feistel cipher.
In a Feistel cipher, $2w$ input bits are split into two halves, (L, R).
(L, R) after one round becomes $(R, L + F(L, K))$ where K is the key (subkey, for that round).

Using a stand alone S-box, not a Feistel cipher, requires the S-box to be invertible. AES uses these. Decryption runs the rounds in reverse.


Plaintext (2w bits)
L_0
w bits
w bits
R_0
Round 1
K_1
F
L_1
R_1
Round i
K_i
F
L_i
R_i
Round n
K_n
F
L_n
R_n
L_{n+1}
R_{n+1}
Ciphertext (2w bits)

Differential Attack

Let V_n denote any n-dimensional vector space over $\mathbb{F}_2$.
Consider the S-Box as $f : V_n \longrightarrow V_n$. (or image V_m)
Consider equations $f(x+a) - f(x) = b$
$a =$ input difference
$b =$ output difference
In an ideal cipher, for any given input difference, a, all output differences b should be equally likely (probability $1/2^n$).
In differential cryptanalysis one exploits an output difference which occurs with much higher probability. This means there exist a, b with $f(x+a) - f(x) = b$ having many solutions.
For a function $f : V_n \longrightarrow V_n$, the *Differential Uniformity* of f is defined to be

$$\max_{b,a\neq 0} |\{x \in V_n : f(x+a) - f(x) = b\}|$$

A function $f : V_n \longrightarrow V_n$ is *perfect nonlinear* (PN) on V_n if its Differential Uniformity is 1.

A function $f : V_n \longrightarrow V_n$ is *almost perfect nonlinear* (APN) on V_n if its Differential Uniformity is 2.
So APN functions have optimal resistance to differential cryptanalysis!

Method first published by Biham and Shamir (1990), also Murphy. (Reduced DES from 2^{56} to 2^{47}.)

It turned out it was known to IBM and NSA in 1970s:

"After discussions with NSA, it was decided that disclosure of the design considerations would reveal the technique of differential cryptanalysis, a powerful technique that can be used against many ciphers. This in turn would weaken the competitive advantage the United States enjoy over other countries in the field of cryptography." – Coppersmith 1994 (about 1970s)