暗号モジュールのサイドチャネル攻撃実験データの

交換用標準フォーマット

Waveform data eXchange Format --- WXF_v1.0

第1版

2008年10月10日(金)

横浜国立大学 大学院環境情報研究院　松本研究室
http://ipsr.ynu.ac.jp/

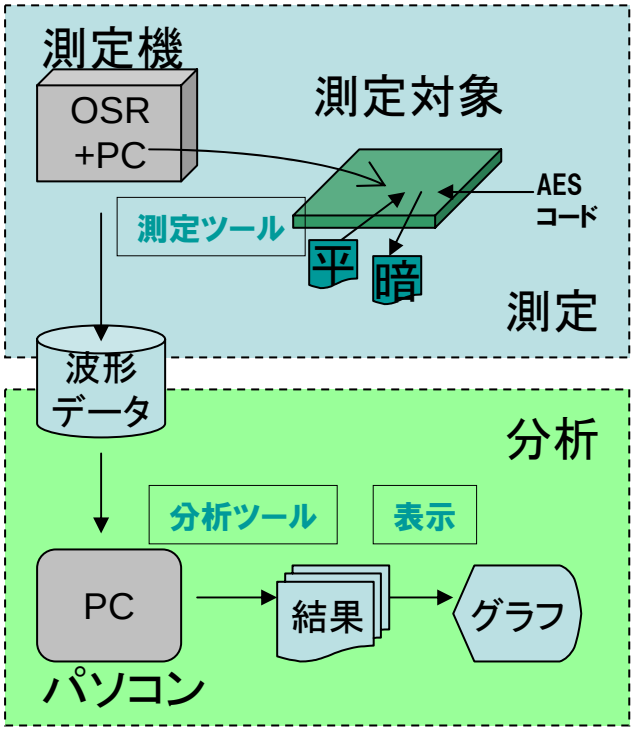
実験データの 交換用標準フォーマット

1. 目的

(1) 主にブロック暗号へのDPA型のサイドチャネル攻撃実験を行うための測定データを対象として，測定データを交換する際のデータ形式(フォーマット)を標準化すること．

(2) 次のことを可能にする：
- 測定ツールは，この標準に従ってデータを整形して出力すれば，少なくともDPA実験に必要な情報(必須部分)を提供でき，その他の情報(オプション部分)を添付してもそれが分析ツールを阻害しないようにできる．
- 分析ツールは，DPA実験に必要な情報を取り込むことができ，必要に応じて，その他の情報を参照できる．

(3) フォーマット標準化によってデータの配布や公開が可能になり，他の環境で測定されたデータを利用できるようになる．新規アルゴリズムの比較検討や，その追試，試験基準の策定に役立てたい．



1.1.「標準化」の考え方

- 既に使用しているデータやツールのうち共通にする価値のある部分を摺り合わせて標準化する．
- 規格として形式的に定義できることよりも実用性や利便性を優先させる．
- 参照できる便利な規格があれば積極的に利用する．市販製品に依存することも避けない．
- 必須部分は最小にして，それ以外は可能な限り利用者が決められるようにする．
- 予約語などで定義されていない部分を任意に追加拡張することを許し，利用状況をみて標準に取り込む．
- データ形式は利用しやすいシンプルなものとし，フォーマットの高度化は様子をみて検討する．

1.2. フォーマットの概要

(1) 波形:
- オシロスコープなどの機器で測定した一区切りの時系列データを「1波形」と呼ぶことにする.
- 1波形を1ファイルに収める(この形式に対応することを必須とする)

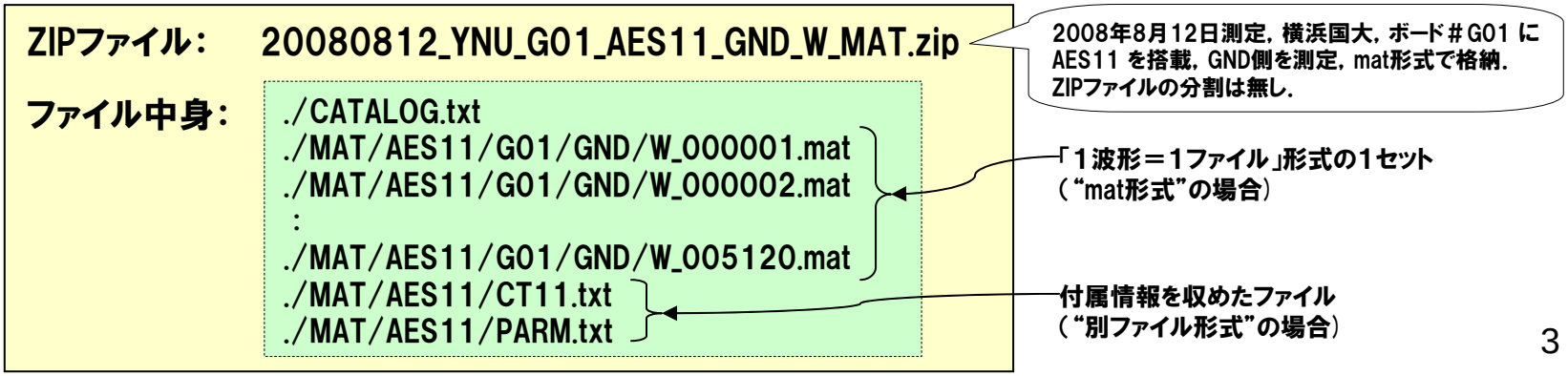
(複数波形＝1ファイル, 1波形＝複数ファイルなどの形式はオプションとし, WXF_v1.0より後のバージョン作成時に検討する)

(2) セット:
- 分析対象となる, ある複数(1個以上)の波形の集まりを「1セット」と呼ぶことにする.
- 測定対象や条件等を1つ定めて, 平文/暗号文を変えて, 収集した一連の波形が1セットになる.
- 2チャンネル以上を使って同時に測定した波形は, 複数セットの波形になる.

(3) 付属情報:
- 平文/暗号文, 鍵, 波形のスケール(垂直, 水平)などの関連する情報をまとめて「付属情報」と呼ぶ.

(4) ZIPファイル:
- 1セットの波形と付属情報をZIPファイルに収める. ZIPファイルは必要により複数個に分割する.
- オプションで, 複数セットを1個のZIPファイルに収めることをサポートする.
- 波形と付属情報を参照する方法をZIPファイル内の "CATALOG.txt" ファイルに記述する.

1.3. ZIPファイルの例:



2. ZIPファイルについて

(1) ZIPファイルの中には，次の3種類のファイルがある．
- 1) "CATALOG.txt"：カタログ情報．必須ファイルでZIP中に1個のみ．
- 2) 波形のセット　：1セット以上を含む．CATALOG.txt で指定される．
- 3) 付属情報　：1組以下．CATALOG.txt で指定される．

- 1ZIPファイルに1セットの波形データを入れることを標準とする．
  オプションで，1ZIPファイルに複数セットを入れることを可能にする．
- 2),3) のファイルは，ボード番号や測定条件等を反映させた適当なサブディレクトリに入れてよい．
- 1ディレクトリのファイル数は数千個程度にして，必要ならばサブディレクトリに分割する．

(2) ZIPファイルのサイズは，数百Mbyte程度を目安にして，必要なら分割する．
- 分割はZIPファイル内のファイルを適当に振り分けることで行う．
- 分割した場合には，ZIPファイル名に「通番of分割数」のような情報を追加する．

(3) ZIPファイルの名前は，人間がみて分かる程度の情報を反映させる．
- 「日付+組織名+中身が分かる文字列+分割情報.zip」のようにする．
  - 例: 20080812_YNU_G01_AES11_GND_W_MAT_(1of4).zip　　非推奨：hoge.zip

（OSやファイルシステムが許容する文字にも注意，Windowsの場合 /¥*|<>:" は使用不可）

2.1. CATALOG.txt について

(1) CATALOG.txt は，ZIPファイル内の情報を示す必須ファイルとする．
- ZIPファイル中のルート直下におく．分割時は1番目に入れる．
- CATALOG.txt内の記述（コメントなど）は英語を使用する．

(2) 情報は，TAG＆VAL形式で示す．TAGは別ページの一覧参照．
- 主な内容：　フォーマットのバージョン，波形のセットのファイル名（ディレクトリ含），
  波形ファイルの構成に関する情報，共通/波形個別の付属情報のファイル名など

2.2. 波形データを入れたファイルについて

(1) ファイル名は，“W_nnnnnn.fff” にする．
- 先頭2文字 “W_” は予約文字で，1波形を収めたファイルを示す．
  「1波形=1ファイル」以外の形式のファイルには，“W”以外の文字を割り当てる．
- nnnnnn は波形の番号（インデックス）を示す6桁の整数．
  1セットの中ではユニークに付ける．複数セットの場合には同時刻に測定したデータは同じ番号にする．
  000000～999999までの100万個を扱える．
- fff は文字列で，波形のデータ形式を示す拡張子．1セットの波形は同じデータ形式とする．

(2) データ形式を示す拡張子は，次の3つを予約する．詳細は別項を参照．
- 1) “mat”: matlab独自形式のバイナリファイル．
- 2) “csv”: csv形式で保存したテキストファイル．
- 3) “bin”: 8bit or 16bit or 64bitの数値を並べたバイナリファイル．

(3) 数値には整数と実数の2種類がある．
- 1) 整数値はオシロのA/D変換直後の値で，
  符号付/符号無の8/16bitの整数(int8/uint8/int16/uint16)型のどれかで表す．
  int8型の場合は-128～127の範囲になる．
- 2) 実数値は1)を実際の測定値に変換(GAIN*data−OFFSET)した値で，
  倍精度64bitの浮動小数点数(double)型で表す．

2.3. 付属情報について

(1) 付属情報は，主に名前(TAG,IDX) & 値(VAL)形式で表現する.
- 付属情報はテキストで表し，CRLFで区切った1行に1情報を入れ，":" をセパレータとし，
  前半のTAGまたはIDX部と，後半のVAL部からなる. mat形式ファイルの中に記載する場合は別途定める.
- TAG部は，数文字程度の英数字で大文字・小文字は同一視する. 別ページに予約TAGを一覧する.
- IDX部は，波形ファイルのファイル名に含まれる数字6桁(インデックス)に対応する.
- VAL部は ":" の次のスペース以外の最初の文字から，スペース以外の最後の文字までとする.
- VAL部は，文字列で半角の英数字・記号のみを使用する(":" と "," と ""(ダブルクォーテーション)は除く).
  VAL部がファイル名を示す場合，OSによって使用不可な記号("¥"等)がある. 使用しないこと.
- "%" で始まる行はコメント行とする. コメント行には半角の英数字・記号のみを使用する.

(2) 付属情報を記載する場所は次の3箇所ある（記述場所に自由度がある）
- A) CATALOG.txt の中に直接.
- B) CATALOG.txt で指定したファイルの中. 1セット共通な情報と波形毎に個別な情報がある.
  - B1) セット共通 ： CATALOG.txtの"T"で始まるTAGで指定. TAG&VAL形式.
  - B2) 波形個別 ： CATALOG.txtの"I"で始まるTAGで指定. IDX＆VAL形式.
  ("MD5SUM", "SHA1SUM" の2つのTAGで指定するファイルは，名前＆値形式とは違う形式になる)
- C) 波形ファイルの中. 波形ファイルの中で記載できるのはmat形式とcsv形式である.
  csv形式の付属情報の行数をCATALOG.txtの"DTLN"で指定する.

(3) 付属情報が重複して記述された場合のデフォルトの優先順位，その他について次のように定める.
- 分析ツールは:
  - A),B1),B2),C) の順番で最後に指定された情報を採用することにする.
    - 例: B1) と C) で同じ情報が記載された場合，C) を採用する.
  - 同一ファイル内で同じTAGが複数記載された場合，最後に出現した情報を採用する.
  - 予約TAGは使用可能な場所が決められている. 他の場所に出現した場合にはエラーにすべき.
    - 例: A) 用のTAGである "FRMT" がB) に出現した場合.
  - 予約TAG以外のTAGが出現した場合には，読み飛ばしてもよいものとする.

3. 予約TAGの一覧

・以下の一覧で “]” で囲まれたTAGはオプションである．“%” で始まる行はコメントである．
・予約TAGを別の意味で使用することは禁止するが，予約TAG以外を使用することは許す．

A. CATALOG.txt のTAG一覧

FRMT	**本フォーマットの名前とバージョンを示す文字列** （WXF_v1.0）		
% **波形データのセットの指定**			
DATA	**1セット分の波形のディレクトリ名とファイル名を示す，** **ファイル名の“*”はインデックスを示し6桁の数字が入る．** **ディレクトリ名の”*”は3桁の数字とする．データ形式は拡張子で指定．**		**例**: MAT/AES11/G01/GND/W_*.mat
[DATA2]	**2セット目，複数セットを含む場合に使用する，**		**例**: MAT/AES11/G01/VCC/W_*.mat
[DATA3]	**3セット目**		
[DATA4]	**4セット目**		
% **波形データに関する情報（直接値を指定）**			
[DTSZ]	**波形ファイルのサンプル数(レコード長）**		**例**: 2500
[DTLN]	**波形ファイルの付属情報の行数(csv形式用）**	**省略時** 0	**例**: 0
[DTBT]	**波形ファイルの数値の形式，** int8, uint8, int16, uint8, double **省略時** mat: double, csv, bin: int8		
% **付属情報ファイルの指定**			
[IPT]	**平文を収めたファイルを示す．**	IDX&VAL**形式**	**例**: MAT/AES11/PT_ONLY11.txt
[ICT]	**暗号文を収めたファイルを示す．**	IDX&VAL**形式**	**例**: MAT/AES11/CT_ONLY11.txt
[IPTCT]	**平文・暗号文を収めたファイルを示す．**	IDX&VAL**形式**	**例**: MAT/AES11/CT11.txt
[ISEND]	**送信データを収めたファイルを示す．**	IDX&VAL**形式**	
[IRECV]	**受信データを収めたファイルを示す．**	IDX&VAL**形式**	
[TPARM]	**セット1共通の付属情報を収めたファイルを示す．**	TAG&VAL**形式**	**例**: MAT/AES11/PARM.txt
[TPARM2]	**セット2用**	TAG&VAL**形式**	
[TPARM3]	**セット3用**	TAG&VAL**形式**	
[TPARM4]	**セット4用**	TAG&VAL**形式**	
[IMemo]	**セット1の波形個別の付属情報ファイルを示す．**	IDX&VAL**形式**	**例**: MAT/AES11/G01/GND/WMEMO11.txt
[IMemo2]	**セット2用**	IDX&VAL**形式**	
[IMemo3]	**セット3用**	IDX&VAL**形式**	
[IMemo4]	**セット4用**	IDX&VAL**形式**	
[MD5SUM]	**md5sumコマンドの出力（MD5によるハッシュ値）を収めたファイルを示す．**		
[SHA1SUM]	**sha1sumコマンドの出力（SHA1によるハッシュ値）を収めたファイルを示す．**		
[README]	**実験データ自身の説明，実験内容，履歴，他データとの関係，波形の測定範囲などの説明を形式フリーで記載したファイルを示す．**		

（3. TAGの一覧のつづき）

B1. セット共通の付属情報を入れるファイル用のTAG一覧

% オシロ関係			
[Instrument]	測定機メーカと型番		例: LeCroy WaveRunner 6040A
[Channel]	測定したチャネル番号	省略時 なし	例: Ch1
[Probe]	使用したプローブ		例: LeCroy PP005A Passive Probe
[VUnit]	縦軸の単位(Gain,Offset適用後の単位)	省略時 V（電圧）	例: uA
[Gain]	測定値への変換用,	省略時 1（変換後のとき）	例: 2.5198744493536651e-002
[Offset]	測定値への変換用,	省略時 0（変換後のとき）	例: 2.5198744493536651e-002
[GainOrig]	スケール変換値(オリジナルの値),	省略時 なし	例: 2.5198744493536651e-002
[OffsetOrig]	スケール変換値(オリジナルの値),	省略時 なし	例: 2.5198744493536651e-002
[SamplingRate]	横軸の単位(サンプリングレート)	省略時 1	例: 5000000000
[DTSZ]	波形ファイルのサンプル数(レコード長)		例: 2500
[DTBT]	波形ファイルの数値の形式, int8, uint8, int16, uint8, double		
		省略時 mat, csv: double64,. bin: int8	
% 測定対象			
[TargetSpec]	ターゲットの名前		
[TargetNo]	ターゲットの個体番号		
[TargetClk]	ターゲットの(暗号ルーチンの)動作クロック		例: 24MHz
[CipherAlg]	搭載した暗号アルゴリズムに関する説明		例: AES
[CipherOpe]	暗号化/復号などの種別を示す	省略時 ENC	例: DEC
[CipherKey]	鍵の値(単純にバイト列で鍵を記述できるとき用)		例: 0123456789abcdef0123456789abcdef
[SendTmpl]	送信データのテンプレート		例: $pt $date $memo
[RecvTmpl]	受信データのテンプレート, 表現方法は次ページを参照		例: $ct
% 測定環境などのメモ類			
[Date]	測定日時		例: 2008-08-22/09:56:01.123
[Temperature]	温度(気温？室温？チップ表面？)		
[Operator]	測定機器を操作した人の氏名		
[Memo]	備考		

C. 波形ファイル中に付属情報を入れる場合のTAG一覧

% B1用のTAGに加えて次のTAGを用意する	
[SEND]	送信データ(中身はSendTmplによって示す)
[RECV]	受信データ(中身はRecvTmplによって示す)

4. IDX&VAL形式のファイルのVAL部分の説明.

B2. 波形ファイル個別の付属情報

```
IPT用        nnnnnn: $pt              $pt は平文を示す. 16進表示.
ICT用        nnnnnn: $ct              $ct は暗号文を示す. 16進表示.
IPTCT用      nnnnnn: $pt $ct          $pt は平文, $ct は暗号文を示す. 16進表示
% ISEND, IRECVのVAL部は, SendTmpl,RecvTmplに従う.
IMemo用      nnnnnn: $memo            $memo はメモを示す. 文字列.
```

・$pt(平文)・$ct(暗号文)などのパラメータは，暗号アルゴリズムや実装によって異なるため全てを決められない．
まずはブロック暗号を想定して，平文・暗号を用意した（他は，今後検討する）.

・IDX&VAL形式のファイルを追加したい場合の拡張方法の例：

・波形毎に"Date"が違う場合には，"I"+"Date" のようなTAGを独自追加する.
CATALOG.txtのTAGで，B2（IDX&VAL形式）のファイルを示すものは Iで始める.

・CATALOG.txt に
```
  IDate: MAT/AES11/TIME.txt
```
のような行を追加する.

・TIME.txt ファイルに,
```
  000001: 2008-08-22 09:56:01
  000002: 2008-08-22 09:56:03
  000003: 2008-08-22 09:56:05
   :
```
と記述する.

実験データの 交換用標準フォーマット

5. データ形式の詳細について

・次の3形式を予約する.

・mat形式：
- ・変数名 w のベクトルデータを，matlabのsaveコマンドでファイル出力したmatlab独自形式のバイナリ.
- ・付属情報を記載するために struct array 型の変数 s をmat形式のファイルに入れてもよい.
  その場合，TAG部をフィールド名に，VAL部をその値に対応させる.
- ・倍精度浮動小数点（8byte）でも圧縮されるためファイルサイズはbin形式と同程度になる．Octaveでも利用可能，
- ・参考：サセボ個体差調査[※1]のデータは，１波形が48kbyte程になる（ZIPしてもサイズはほぼ同じ）.

・csv形式：
- ・CRLFで区切られた１行に１サンプルの値を，整数あるいは浮動小数点表示でファイル出力したプレーンテキスト.
- ・オプションでファイルの先頭に付属情報を記載可能とする.
  （matlabでは save –ascii –doubleで出力，txtreadコマンドで指定行数をスキップして読み込める）
- ・1レコード（=1行）には１フィールドのみがあり（","を含まない），CSVとしてのヘッダ行は無いものとする（RFC4180を参照）
- ・フィールドは文字列であっても，ダブルコーテーションで括らない．最後のレコードにもCRLFを付ける.
- ・double型のデータをファイル出力するとサイズが大きめ（10倍以上）になり，HDDに展開するには具合が悪い.
  - ・ZIP圧縮したまま使用するならば問題はない.
- ・参考：サセボ個体差調査のデータは，double型だと1319kbyteで，ZIPすると61kbyte程になる.

・bin形式：
- ・8bit or 16bit or 64bitのデータを，ヘッダや構造無しに単純に並べたバイナリファイル.
  - ・8bitは int8型 または uint8型とする.
  - ・16bitは int16型 または uint16型で，Low byteを先にする（little-endian）
  - ・64bitは double型で，IEEE754のlittle-endianとする.
- ・オシロから取り出したデータを，OffsetやGainなどのスケール変換をせずに保存した値を入れることを想定.
- ・参考：サセボ個体差調査のデータは，8bit長×50kSなので，50kbyteのファイルになる.

※1）本資料で，サセボ個体差調査のデータとは，http://ipsr.ynu.ac.jp/ssb/index.html のデータのことを指す.

X. 補足情報

X.1. ZIPファイルのシンプルな例（mat形式で交換する場合）

・測定データそのものである matファイルのほかには，3つのテキストを同封すればよい．

ZIPファイルの中身

```
./CATALOG.txt
./MAT/AES11/G01/GND/W_000001.mat
./MAT/AES11/G01/GND/W_000002.mat
 :
./MAT/AES11/G01/GND/W_005120.mat
./MAT/AES11/CT11.txt
./MAT/AES11/PARM.txt
```

CATALOG.txt の中身

```
FRMT: WXF_v1.0
DATA: MAT/AES11/G01/GND/W_*.mat
IPTCT: MAT/AES11/CT11.txt
TPARM: MAT/AES11/PARM.txt
```

・このファイルは必須．
・ZIPファイル内の他のファイルの名前を示す

PARM.txt の中身

```
SamplingRate: 5000000
CipherAlg: AES
```

・必須項目は無いので省略も可能

CT11.txt の中身

```
000001: 00000000000000000000000000000001 646da25c4e6ffb690f9cfe1f59b2e0d6
000002: 00000000000000000000000000000002 f7c6454d35bfdd32940a21bb25662b29
000003: 00000000000000000000000000000003 59bb09a39a8af6e8cd2f4ee51ff70cd0
000004: 00000000000000000000000000000004 5758b14de35d235eaf2bc3e47611095a
（以下，5120まで続く）
```

・csv形式でファイル先頭部分に平文・暗号文を付けるならば，このファイルは同封しなくてもよい．
・その場合は CATALOG.txt で付属情報の行数 DTLN を指定する．

X.2. オシロのバイナリファイルからmatlabへの読み込みと出力

・次の3社のオシロのバイナリデータは，matlabで読み込むためのスクリプトが公開されている．

Agilent：	importAgilentBin.m （Agilent社のWEBでも配布されている）	
LeCroy：	ReadBinaryWaveforms.m	
Tektronix：	wfm2read.m	参照 http://www.mathworks.com/matlabcentral/

・型変換，matlab上ではdoubleになっているので適当な型に変換する（例：int8型）．

```
w_int8=int8((w+OffsetOrig)/GainOrig);
```

・mat形式の出力： `save('W_123456.mat','-MAT','w'); %% double`のとき
・csv形式の出力： `save('W_123456.csv','-ASCII','w_int8'); %% int8`のとき
`save('W_123456.csv','-ASCII','-DOUBLE','w'); %% double`のとき
・bin形式の出力： `fp=fopen('W_123456.bin','w','l'); fwrite(fp,w_int8,'int8'); %%int8`のとき
`fp=fopen('W_123456.bin','w','l'); fwrite(fp,w,'double'); %%double`のとき

X.3. md5sumの出力例を示す．sha1sumの出力も同様である．

```
$ md5sum MAT/AES11/G01/GND/W_*.mat MAT/AES11/*.txt
4dc924f51df76fb8d9e46a814a6cb2ec *MAT/AES11/G01/GND/W_0001.mat
078af2c081b2114825848c4205693ee5 *MAT/AES11/G01/GND/W_0002.mat
5934f42e8ac1566e6cc2582d1cb8c6a2 *MAT/AES11/G01/GND/W_0004.mat
93fac4086608233c3b5ab3e5eb7ea290 *MAT/AES11/G01/GND/W_0005.mat
（途中省略）
8c28e7f824afff45ade9884c08e0a8ab *MAT/AES11/CT11.txt
05723e22523fab7dbde785456e62bf35 *MAT/AES11/PARM.txt
$
```

コマンド行

出力

X.4. mat形式のファイルの中に入れた付属情報の例

・下記にmat形式のファイル中に付属情報を幾つか入力した例を示す.

```
>> s
s =
       Instrument: 'LeCroy WaveRunner 6040A'
          Channel: 'Ch1'
            Probe: 'LeCroy PP005A Passive Probe'
            VUnit: 'uA'
             Gain: 0.0252
     SamplingRate: 5.0000e+009
             SEND: 'B9D0C38E348EE770F9454A77A177FA07'
             RECV: '501F2D4983666EB07DCCE771F558D6BC'
>>
```

Y. 検討課題

Y.1. WXF_v1.0より後のバージョン作成時等での検討課題

(1) TAGの拡張性(XML化の要否)
(2) ブロック暗号以外のTAGの追加, 値(VAL)で使用する単語やフレーズの整理.
(3)「複数波形=1ファイル」の検討, どのような波形を1ファイルにいれるか整理が必要.

Y.2. 波形データフォーマットを超えた課題:

・現状での制約事項: 選択平文などインタラクティブな分析は実験データを交換することでは追試できない, にどう対処するか.

Z. 本文書作成の経緯と履歴情報

・この文書は, 横浜国立大学大学院環境情報研究院 松本研究室のハードウェアセキュリティチームが, 2007年度に同チームが実施したサセボ個体差調査のデータ配布用ファイル形式を原型として, サイドチャネル攻撃実験データの交換に広く適用できるフォーマットWXFを定めるために作成した.
・フォーマットWXFの検討に際しては, CRYPTREC暗号モジュール委員会電力解析実験ワーキンググループ(WG)の委員および事務局各位から多くの有益なご指摘ご意見を頂戴した. 特に同WG委員の黒川防衛大学校教授の研究室には初期の段階でご討論を戴いた. ここに検討に加わって戴いた皆様に感謝の意を表したい. ただし, 本文書およびフォーマットWXF_v1.0に潜在する誤りがあるとすればその責任は全て松本研究室 責任者(松本 勉)に帰するものである.

・本文書のお問合せ先: URL http://ipsr.ynu.ac.jp/wxf/index.html

Z.1. 履歴情報

・2008年8月12日 (火) 初版作成
・2008年8月22日 (金) 防衛大学校, IPAとのお打合せ版
・2008年9月01日 (月) CRYPTREC暗号モジュール委員会 電力解析実験ワーキンググループ(WG)事務局とのお打合せ
・2008年9月03日 (水) WGでの説明用
・2008年9月24日 (水) WGコメントの反映の検討版
・2008年9月30日 (火) WG事務局とのお打合せ版
・2008年10月01日 (水) WGコメント反映版
・2008年10月03日 (金) WGコメント反映版(誤記修正)
・2008年10月10日 (金) 第1版制定

End of Page