CyberSecurity: Principle and Practice

BSc Degree in Computer Science 2022-2023

Lesson 3: Encoding

Prof. Mauro Conti

Department of Mathematics University of Padua conti@math.unipd.it http://www.math.unipd.it/~conti/ Teaching Assistants

Pier Paolo Tricomi pierpaolo.tricomi@phd.unipd.it Tommaso Bianchi tommaso.bianchi@phd.unipd.it





Motivations



- The world is full of different "environment"
 - o e.g., nation, music notation
- Each "environment" follows its own standard
 - o e.g., language, rules
- We need interaction strategies

Encoding



Why Encoding data?

- Transform data so that it can be properly (and safely) consumed by a different type of system
- The goal is <u>not</u> to keep information secret
 - e.g., keys are not required
- Encoding can be <u>easily</u> be reversed
 - easy to recognize an encoding strategy
- Decoding is the inverse operation

decimal			binary
\rightarrow	65	\rightarrow	01000001
\rightarrow	66	\rightarrow	01000010
\rightarrow	67	\rightarrow	01000011

Encoding



Some examples of encoding:

- base64
 - web communication
- Hexadecimal
 - simplified version of binary in CS
- Uuencoding
 - binary to text encoding for UNIX

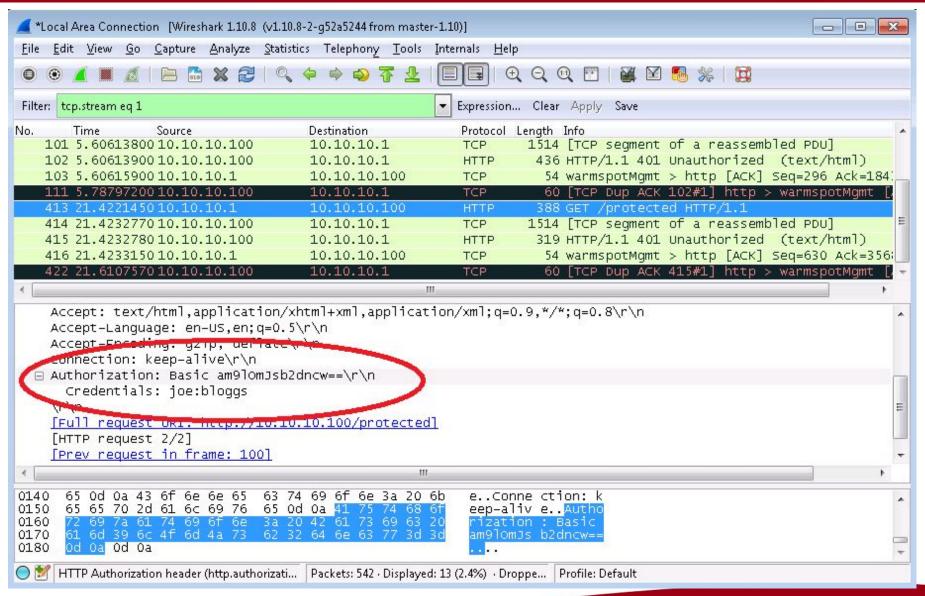
Base64



- Very common in web communications
- Lengths of final messages always a multiple of 4
- Unique alphabet
 - \circ [A-Z, a-z, 0-9, +, /, =]
 - \circ 0 = A, 1 = B, ..., 26 = a, 27 = b, ...
 - \circ 52 = 0, 53 = 1, ..., 62 = +, 63 = /
- Padding
 - Might end with "==" or "="

Base64





Base64



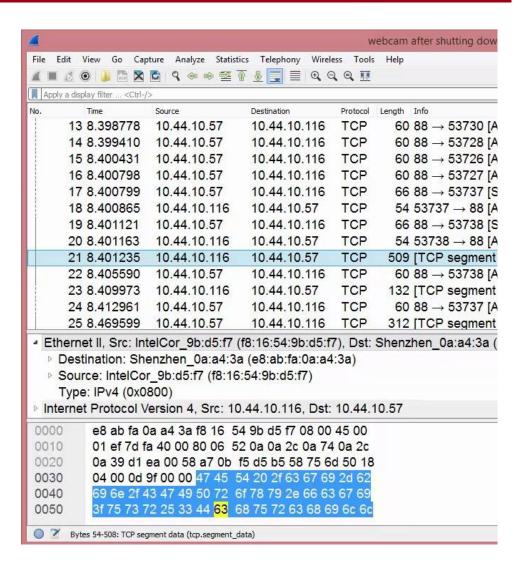
Some examples are:

- pleasure -> cGxlYXN1cmU=
- leasure -> bGVhc3VyZQ==
- easure -> ZWFzdXJI
- asure -> YXN1cmU=
- sure -> c3VyZQ==

Hexadecimal



- Similar to base64
- The alphabet is [A-F, 0-9]
- Widely used to represent
 - MAC Address
 - Memory dumps



UUEncoding



- Always start with <u>begin</u> followed by the <u>mode</u> and the <u>file name</u>
- Always end with both
 - O "'"
 - end

```
begin 600 test.txt
M5&AI<R!I<R!A('1E<W0@9FEL92!F;W(@:6QL=7-T<F%T:6YG('1H92!V87)I
M;W5S"F5N8V]D:6YG(&UE=&A09',N($QE="=S(&UA:V4@=&AI<R!T97AT(&Q0
M;F=E<B!T:&%N"C4W(&)Y=&5S('10('=R87'@;&EN97,@=VET:"!"87-E-C0@
E9&%T82P@=&]0+@I'<F5E=&EN9W,L($9R86YK(%!I;&A09F5R"@''
end
```

Encoding in Practice



- When we analyze data, it might be represented in an unknown encoding
- How to identify the proper encoding?
 - experience
 - alphabet
 - patterns
 - origin of data



Questions? Feedback? Suggestions?







