FernUniversität in Hagen

CUING

Trends in Stegomalware: Techniques and Countermeasures

Wojciech Mazurczyk, Ph.D., D.Sc.

FernUniversität in Hagen, Germany

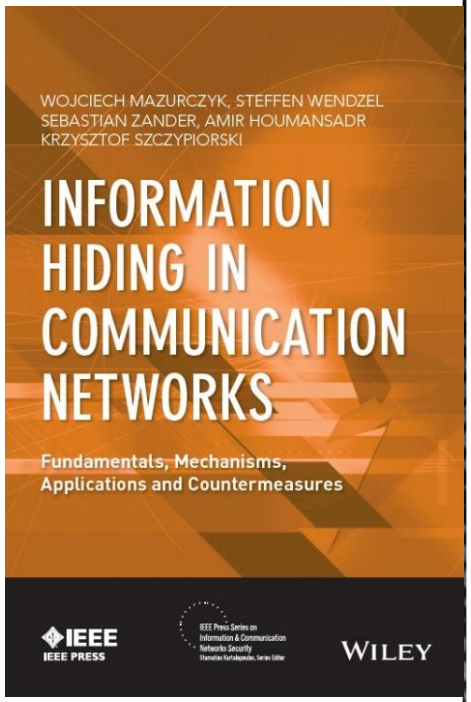
3rd International Conference on Frontiers in Cyber Security (FCS 2020)

SIMARGL

18 December 2020 **Tianjin, P.R. China**

About me

- **Professor** at Institute of Computer Science, Warsaw University of Technology (Poland) and **Senior Researcher** at FernUniversitaet in Hagen (Germany)
- **Head** of the Computer Systems Security Group (CSSG) at WUT
- Author or co-author of 2 books, over 150 papers, 2 patent applications and over 35 invited talks
- **Research interests**: information hiding, network security, bio-inspired security, network traffic measurements
- Involved in many **research projects** funded by: EC (H2020: SIMARGL, PREVISION, IoRL), US Army (CoCoDe), and domestic ones
- Founder and Coordinator of the **Criminal Use of Information Hiding (CUIng)** initiative (with Europol)



SIMARGL

- **SIMARGL (Secure Intelligent Methods for Advanced RecoGnition of malware and stegomalware)**
- The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement **No 833042**
- SIMARGL brings together experts in information hiding and malware from 14 European organisations from 7 countries

CUING

SIMARGL

The SIMARGL solution

Detection

Introduce new and innovative techniques to detect stegomalware, including machine and deep learning methods

Toolkit

Produce a toolkit that enables organisations to easily detect and counter stegomalware

Training

Provide training to Law Enforcement and other end-users to improve awareness of information hiding techniques

Deployment

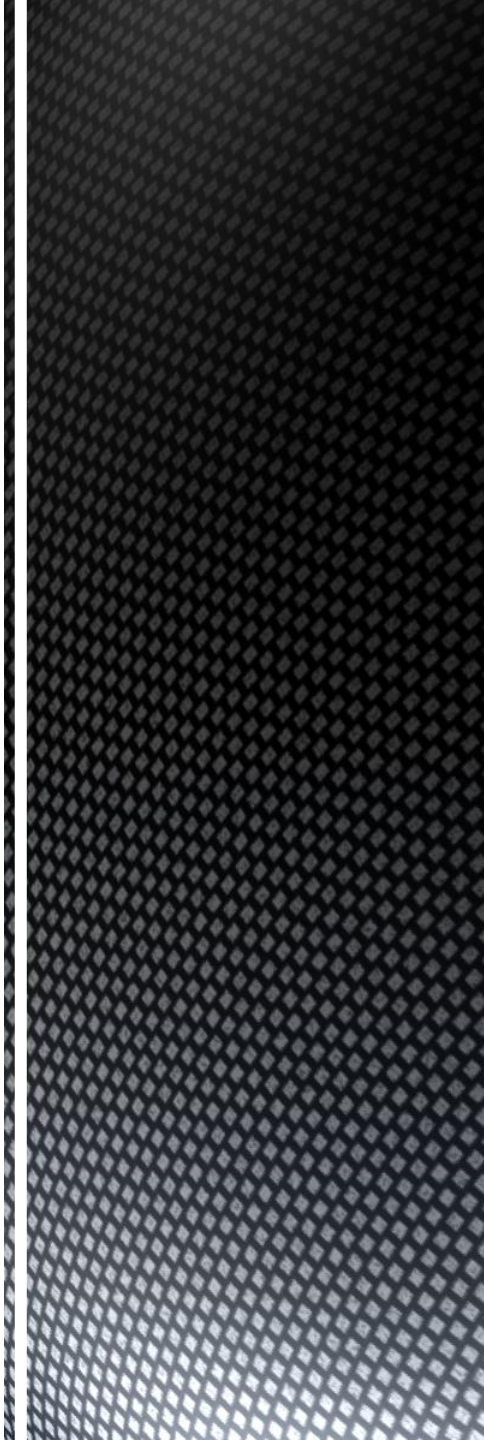
Deploy the SIMARGL results in real world use-cases that enable the approach to be validated

Project website: https://simargl.eu

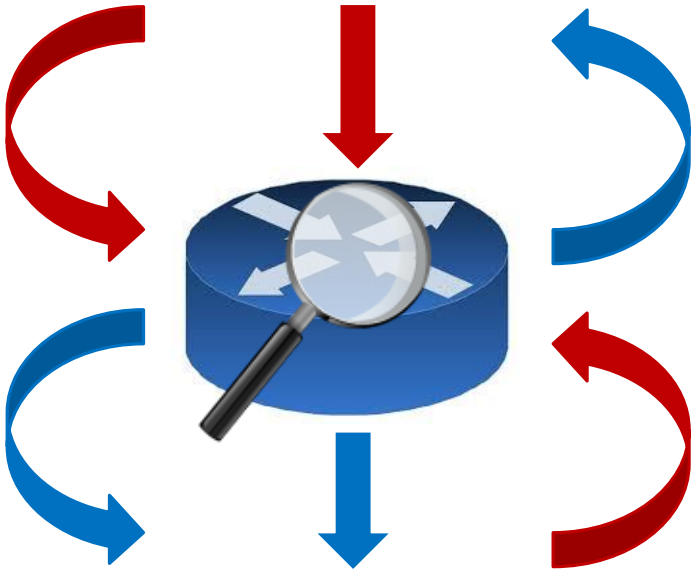
Agenda

- **Introduction** to information hiding
- Information hiding techniques in **real-life malware**
- **Trends** in network covert channels
- **Challenges for countering** network information hiding

Introduction to Information Hiding



Information hiding in networks: an analogy



Information hiding: inspiration

- **Information hiding** is part of a wide spectrum of methods that are used to make secret data difficult to notice for the curious third party observers
- **Steganography** is one of the most well-known subfields of information hiding and aims to cloak secret data in a suitable carrier – in communication networks we use the term **network covert channels** or **network steganography**
- Information hiding has proved very handy and has been **utilized and mastered by humankind throughout the ages**
- Inspiration for such mechanisms is strongly related to phenomena observable in nature as they **have their roots in nature** (camouflage, mimcry, etc.)

What influenced development of modern information hiding?

- Terrorist attack on **11th September 2001** (probably) was planned using steganography
- After these attacks there has been a **growing interest in modern methods** and their detection

02/05/2001 - Updated 05:22 PM ET

USA TODAY

Home
News
Money
Sports
Life
Tech
Main Categories
Tech briefs
Web Guide
Tech Investor
Product reviews
More Tech
Columnists
Shareware Shelf
Talk Today
Weather

ARCHIVES
SEARCH FOR NEWSPAPER ARTICLES
CLICK HERE

NEW E-MAIL
GET NEWS

Tech

▪ E-mail this story ▪ Subscribe to the newspaper ▪ Sign-up for e-mail news

02/05/2001 - Updated 05:22 PM ET

Terrorist instructions hidden online

By Jack Kelley, USA TODAY

WASHINGTON — Osama bin Laden and other Muslim extremists are posting encrypted, or scrambled, photographs and messages on popular Web sites and using them to plan terrorist activities against the United States and its allies, U.S. officials say. The officials say bin Laden and his associates are using the Internet to conduct what some are calling "e-*jihad*," or holy war. Bin Laden, a dissident Saudi businessman, has been indicted for the 1998 bombing of two U.S. embassies in East Africa and is believed to be responsible for last fall's bombing of the USS Cole in Yemen. Four alleged bin Laden associates went on trial Monday in federal court in New York for the embassy bombings. "To a greater and greater degree, terrorist groups, including Hezbollah, Hamas, and bin Laden's al Qaida group, are using computerized files, e-mail, and encryption to support their operations," CIA Director George Tenet wrote last March to the Senate Foreign Relations Committee. The testimony, at a closed-door hearing, was later made public.

Information hiding in use by **real spies**

- **June 2010: a Russian spy ring discovered in US**

BBC Mobile | News | Sport | Weather | Travel | TV

NEWS US & CANADA

Home | UK | Africa | Asia-Pac | Europe | Latin America | Mid-East | South Asia | US & Canada | Business | Healt

29 June 2010 Last updated at 03:17 GMT

FBI allegations against 'Russian spies' in US

Court papers setting out the allegations against 10 people arrested on suspicion of spying for the Russian government reveal details worthy of a Cold War spy novel.

The espionage ring was allegedly trained by the Russian Foreign Intelligence Service (SVR), whose headquarters are known as Moscow Centre.

Its methods are said to have ranged from the hi-tech, such as [illegible] to [illegible]

Ten suspects have been arrested

HIDDEN MESSAGES

Some of the suspects are accused of using steganography - a method of concealing data in an image using special software - to pass information to Moscow Centre by posting pictures on public websites.

Using data and a 27-character password gained by searching a New Jersey property in 2005, US agents accessed a steganography programme that led them to websites where they found certain images, the court documents say.

"These images appear wholly unremarkable to the naked eye. But these images (and others) have been analysed using the steganography program. As a result of this analysis, some of the images have been revealed as containing readable text files."

Indictment act:

(http://www.justice.gov/opa/documents/062810complaint2.pdf)

III. MEANS AND METHODS OF THE CONSPIRACY

A. SECRET COMMUNICATIONS

20. To further the aims of the conspiracy, Moscow Center has arranged for the defendants clandestinely to communicate with the Russian Federation. In particular, the conspirators have used, among others, the secret communications methods described below – steganography and radiograms.

1. STEGANOGRAPHY

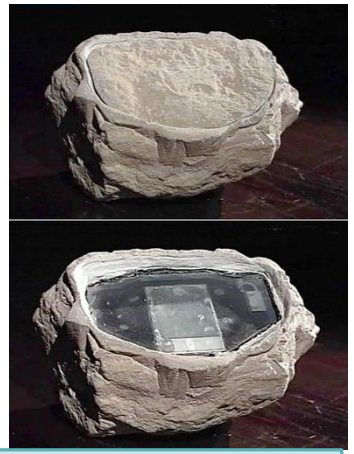
21. Steganography is the process of secreting data in an image. Moscow Center uses steganographic software that is not commercially available. The software package permits the SVR clandestinely to insert encrypted data in images that are located on publicly-available websites without the data being visible. The encrypted data can be removed from the image, and then decrypted, using SVR-provided software. Similarly, SVR-provided software can be used to encrypt data, and then clandestinely to embed the data in images on publicly-available websites.

22. As is set fo[illegible] of the Illegals have communicated with Mosc[illegible] of steganography. In each of the three judici[illegible]sidential searches referenced above (the 2006 [illegible] 2006 Seattle search, and the 2005 New Je[illegible] enforcement agents observed and forensically c[illegible]puter disks ("Password-Protected Disks" [illegible]quent investigation as described below, I belie[illegible]rd-Protected Disks contain a steganography pro[illegible]he SVR and the Illegals.

Information hiding applications

Information hiding includes various techniques that can be broadly divided, based on the aim to be achieved, into two groups:

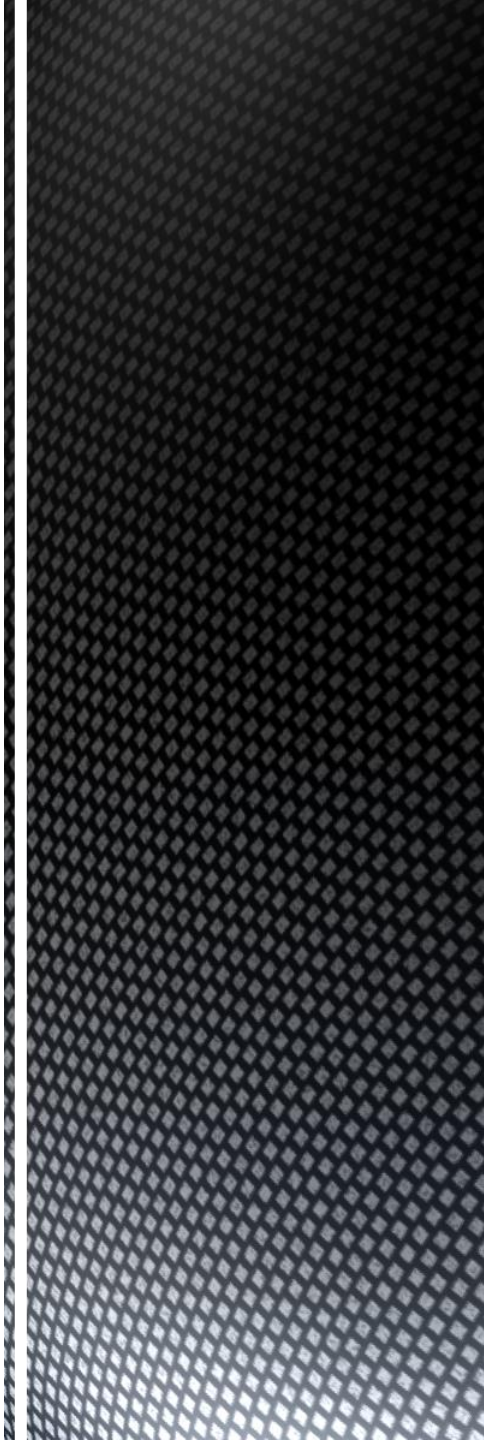
- ***„Safe locker”*:** solutions that allow to hide secret data in such a way that **no one besides the owner is authorized to discover its location and retrieve it**. In other words, the aim is to not reveal the stored secret to any unauthorized party



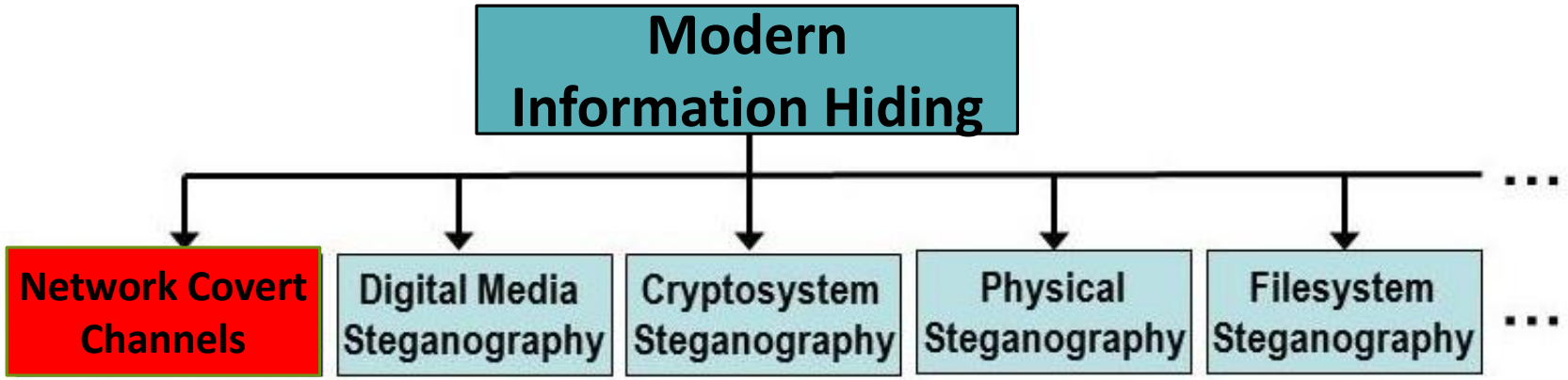
- ***„Live drop”:*** methods intended for the communication of messages with the aim of **keeping some aspect of such an exchange secret**

OUR MAIN FOCUS

Information concealment in **communication networks**



Many types of data hiding techniques

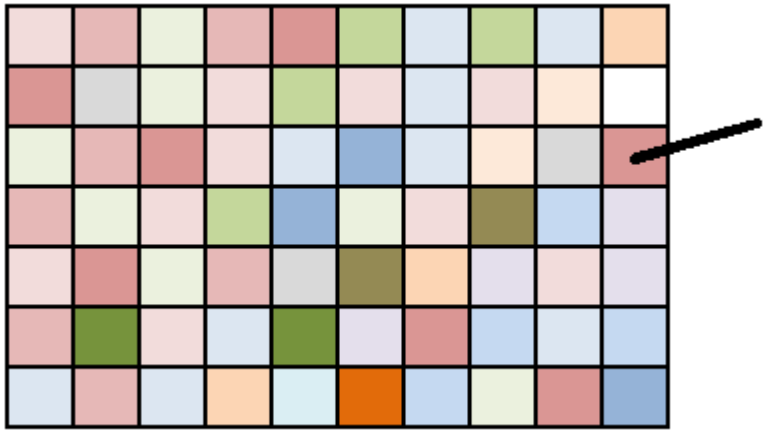


Many possible carriers exist – hard to monitor all of them!

Other types of information hiding are also possible:

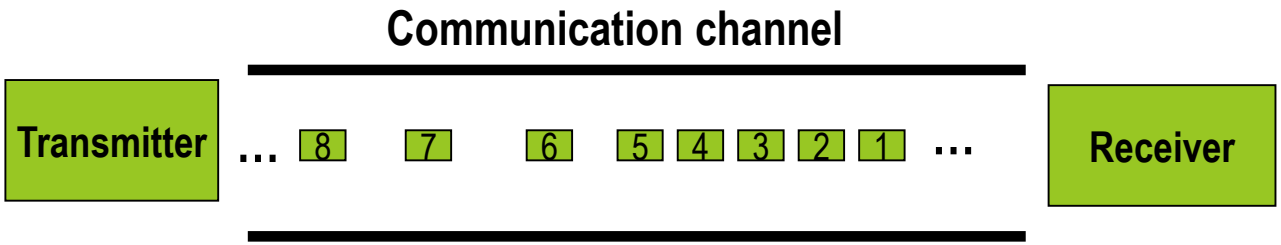
- Network traffic type obfuscation techniques
- Local covert channels
- Etc.

Digital image vs. network traffic



Characteristic features:

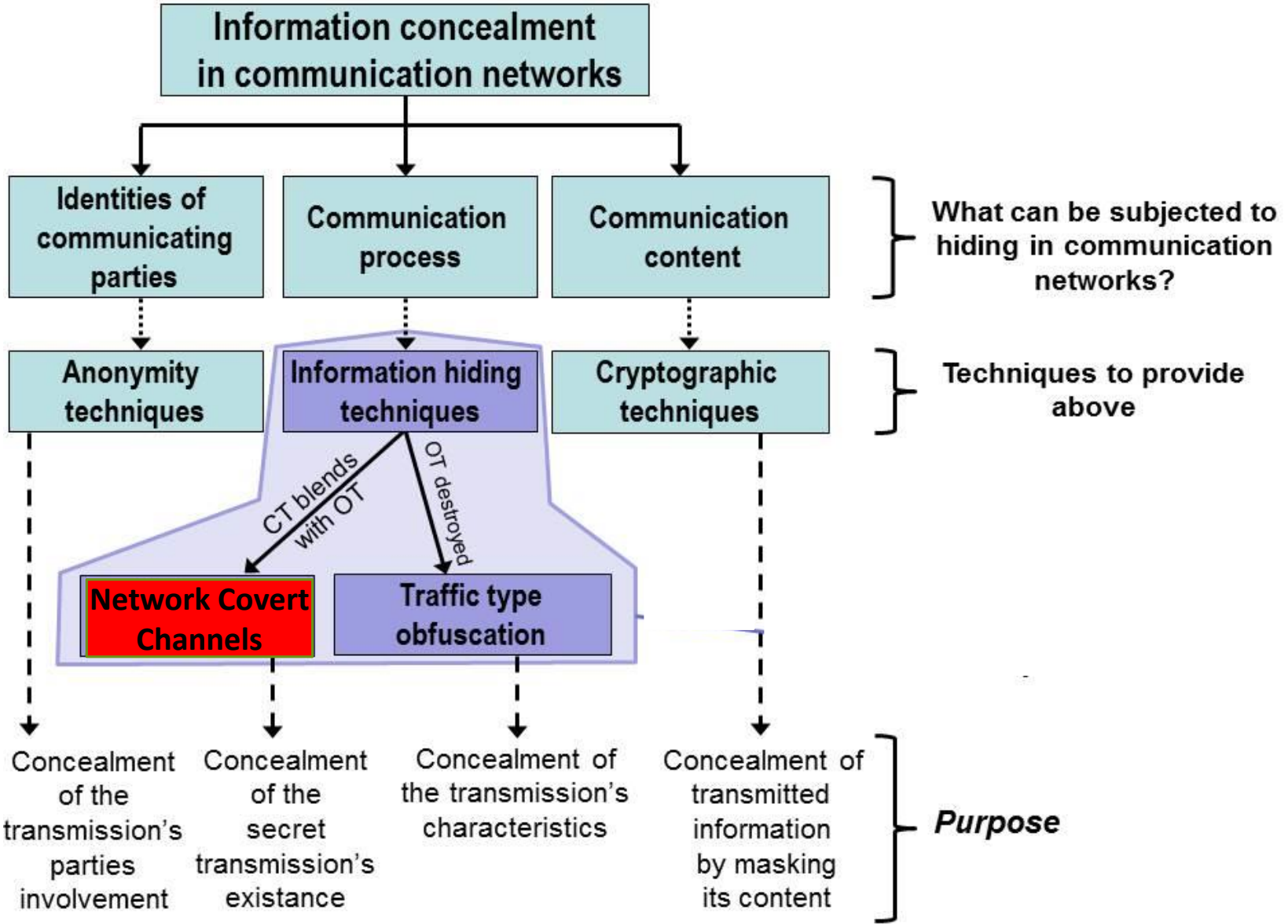
- Limited capacity - known in advance
- Limited data hiding „dimensions”
- Artifacts remains
- Static in nature



Characteristic features:

- Many potential data hiding „dimensions”
- Emphemeral in nature
- Dynamic in nature

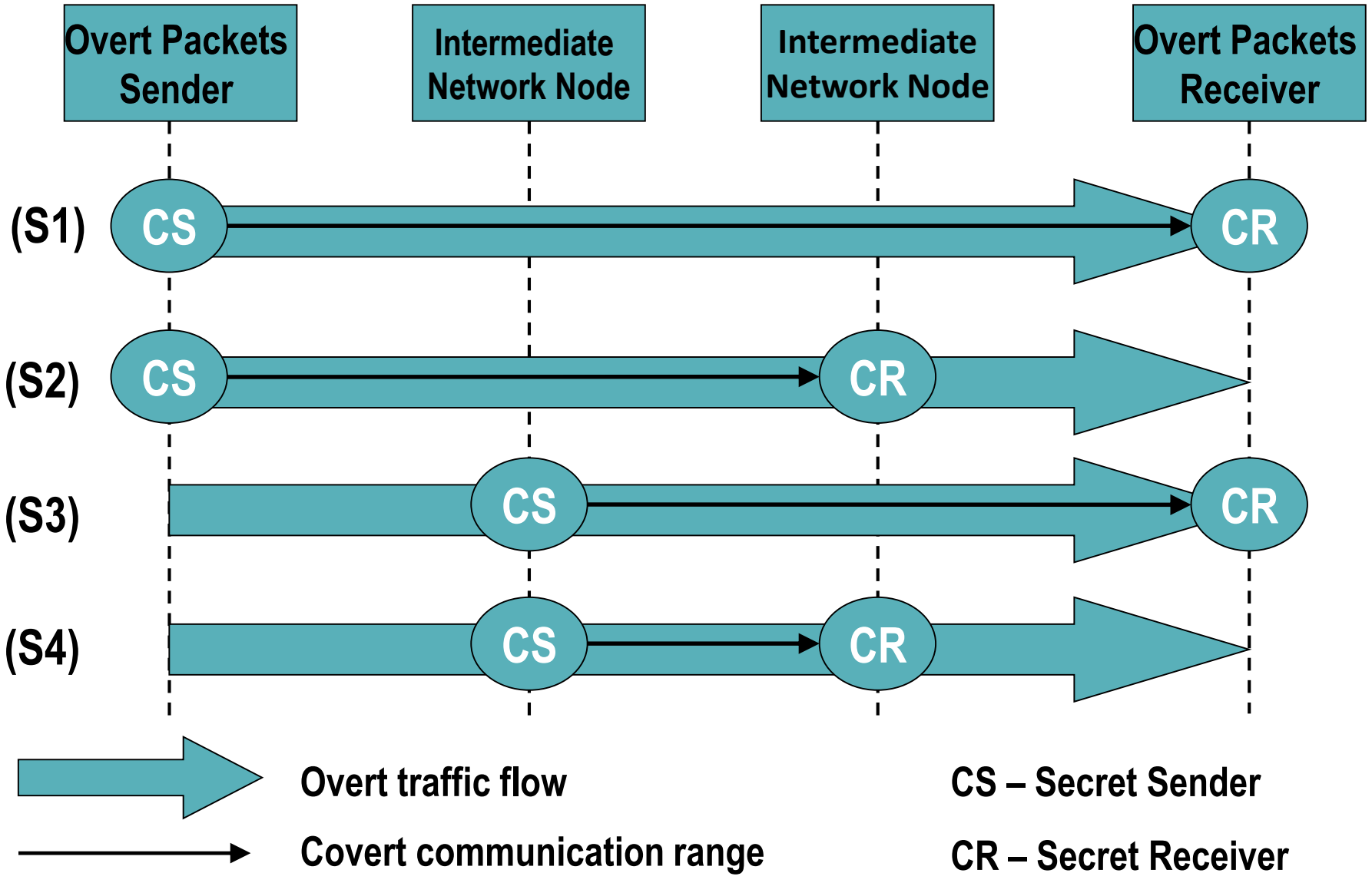
What can be subjected to hiding in communication networks?



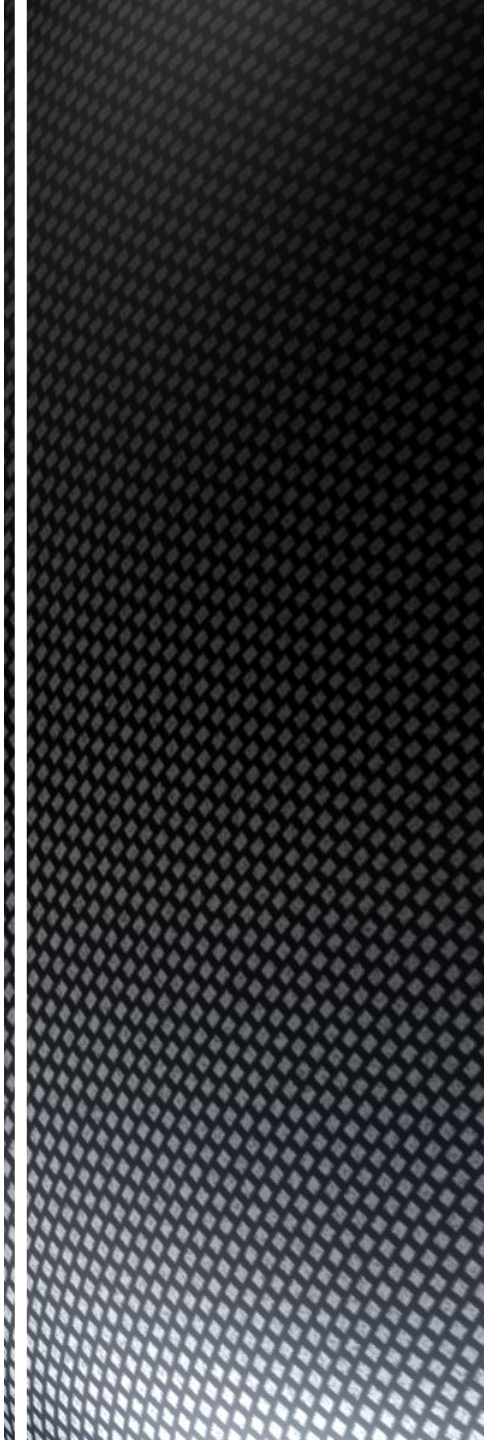
Network Covert Channels

- **One of the newest trend in information hiding**
- Data hiding methods that utilize as a hidden data carrier **network protocols** (PDUs and/or the way they are exchanged) **and/or the relationships between them**
- **Information hiding opportunities** in networks come from the **increasing complexity and redundancy of protocols/services**
- **To provide undetectability:**
  - Use of popular carriers
  - Use of anomalies that happen in the networks
  - Imitate behavior specific to certain types of traffic / protocols / services / users (mimicry)

Hidden communication scenarios



Information hiding techniques in **real-life malware**



Secu

e the

oses

BROUGHT TO YOU BY

SecurityInte

esearch Podcast Ev

S

fo

Nov

FORTI

Partners

87
Shares

Eugene Kaspersky
@e_kaspersky

Forget about #cryptography - #steganography is the new black for cyberespionage and the #PLATINUM APT group is here to use it to fly under #cybersecurity radars kas.pr/platinum via @securelist

Platinum is back

In June 2018, we came across an unusual set of samples spreading throughout South and Southeast Asian countries targeting diplomatic, government and ...

securelist.com

the past several years,
around for some time,

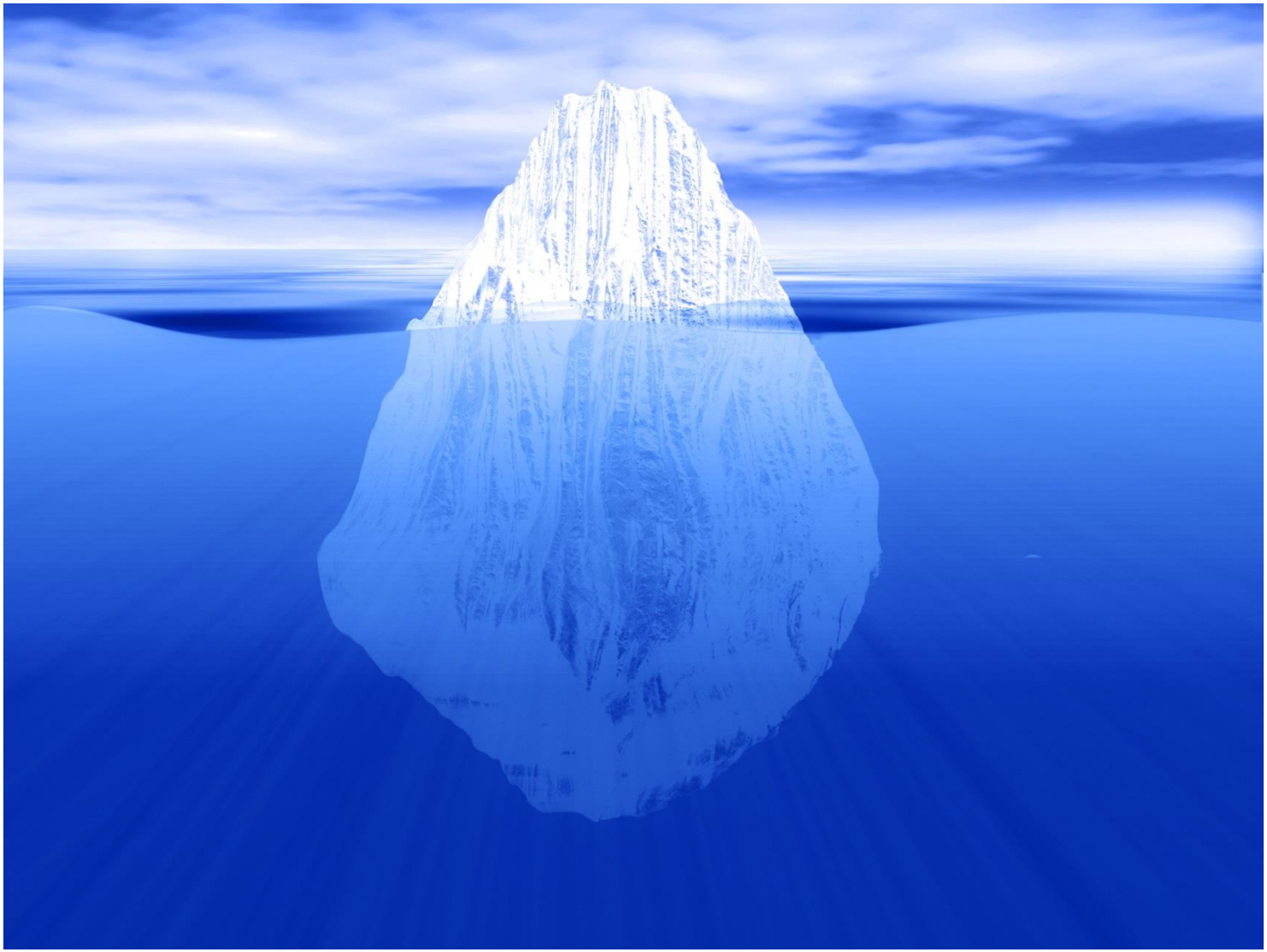
s.

gly becoming a go-to

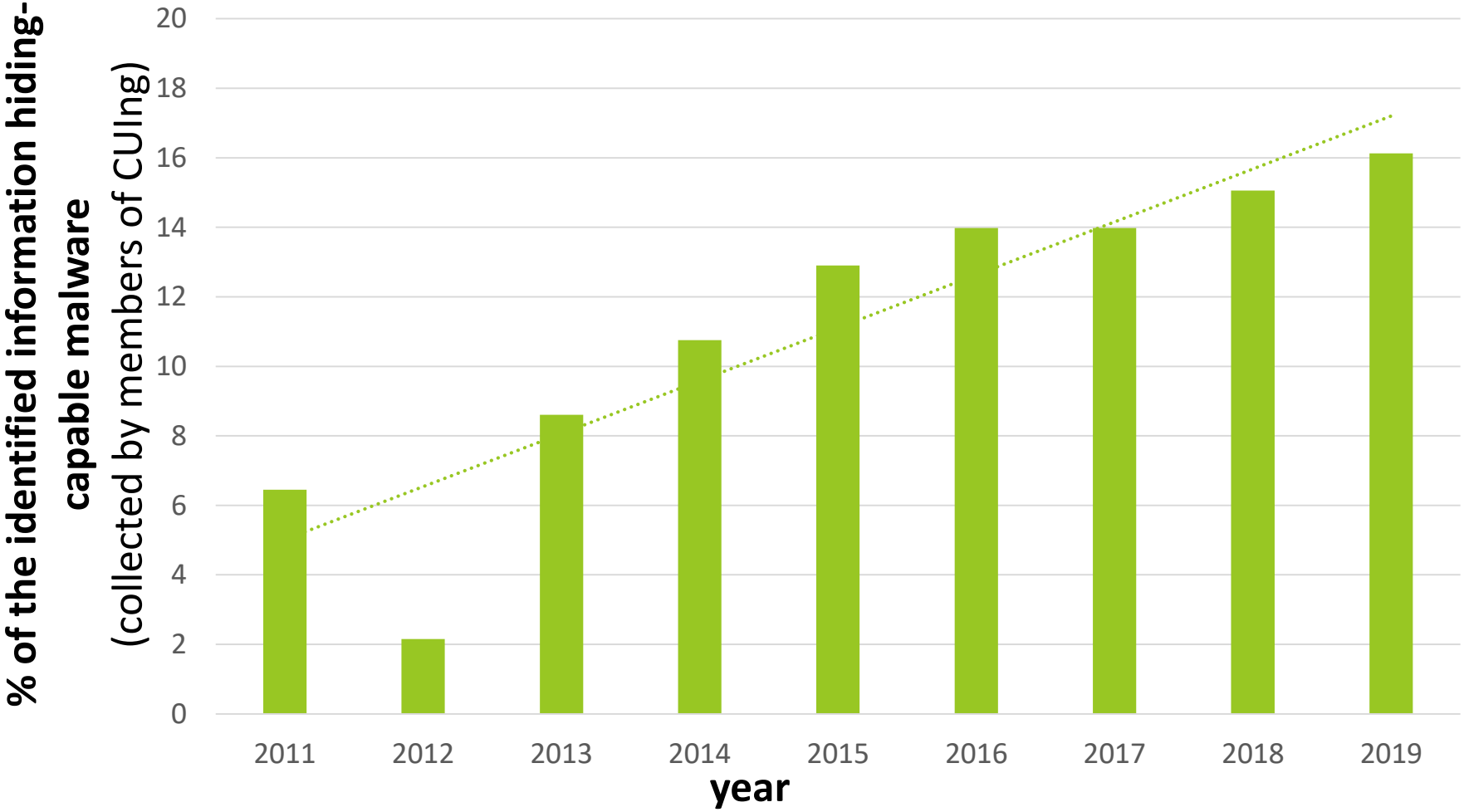
Roots of new trend: Trojan.Downbot

- The first massive usage of such techniques can be traced back to **2006** when **Operation Shady RAT** led to attacks against numerous institutions worldwide and inflicted damage for months
- The main program responsible for this attack was the **Trojan.Downbot**
- This trojan created a back door and then downloaded files appearing as real **HTML pages or JPEG images**
- These files were encoded with commands that would allow **remote servers to gain access to local files** on the infected host computer

How common is utilization of information hiding techniuqes by current malware?



Increase in information-hiding capable malware

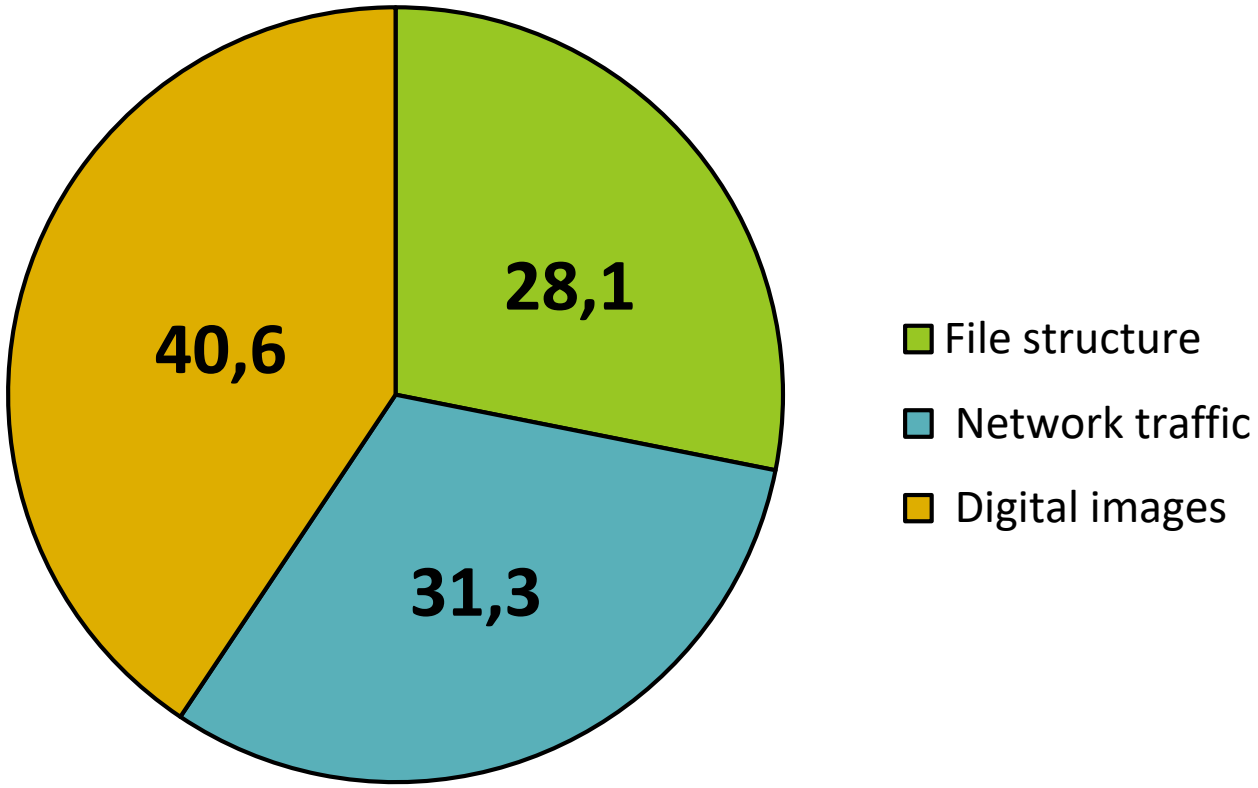


Information hiding-capable malware is **heavily underestimated**: security experts do always not correctly recognize and classify techniques used

Classification of information hiding-capable malware

- ***Group 1:*** malware that embeds secret data by **modifying a digital image file's structure**
- ***Group 2:*** malware that embeds secret data by **using digital media steganography**
- ***Group 3:*** malware that injects secret data **into network traffic**

Distribution of information hiding-capable malware



Malware that embeds secret data by **modifying a digital image file's structure**

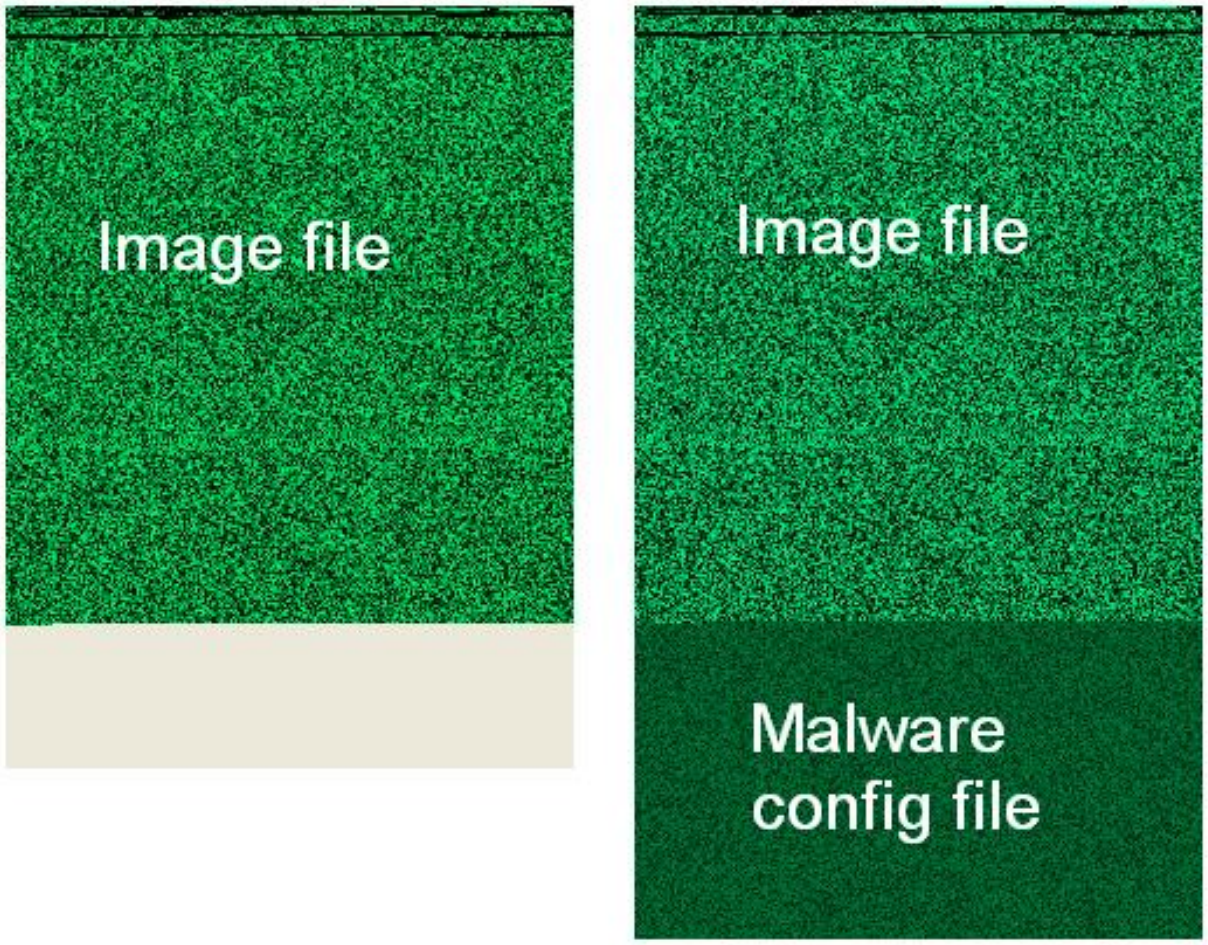
- **ZeusVM** – variant of Zeus/Zbot malware
- Similar principle has been found e.g. in (infamous) **Hammertoss** APT discovered in July 2015
- It downloads innocent JPG from the C&C server:

Address	Value	Comment
01F5F380	01E129A8	ASCII "Mozilla/4.0 (compatible; MSIE 6.0; Windows N
01F5F384	00CC0008	
01F5F388	01716FA4	ASCII "GET"
01F5F38C	01E12A40	ASCII "/prefer/stars/rihannew.jpg"
01F5F390	01714DC8	ASCII "HTTP/1.1"
01F5F394	00000000	
01F5F398	01742530	
01F5F39C	8484F700	
01F5F3A0	00000000	
01F5F3A4	01F5F7C6	ASCII "https://bilance.humanwebcentr.net:63992/pref

- The file shows a sunset

ZeusVM: how to discover the hidden data

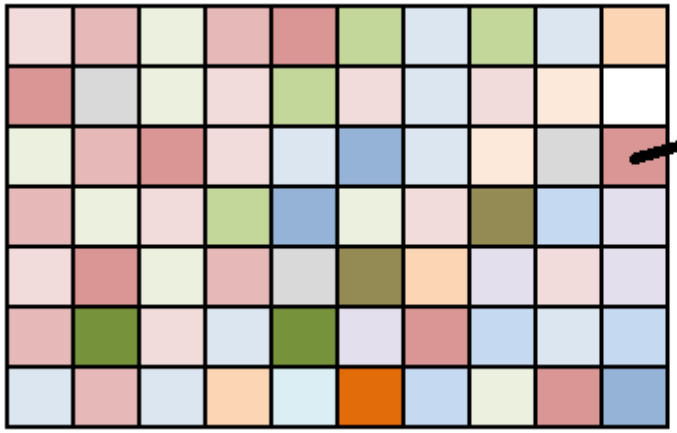
- Malware config file **appended after the original image data** – the image still launches correctly!



Malware that embeds secret data by **using digital media steganography**

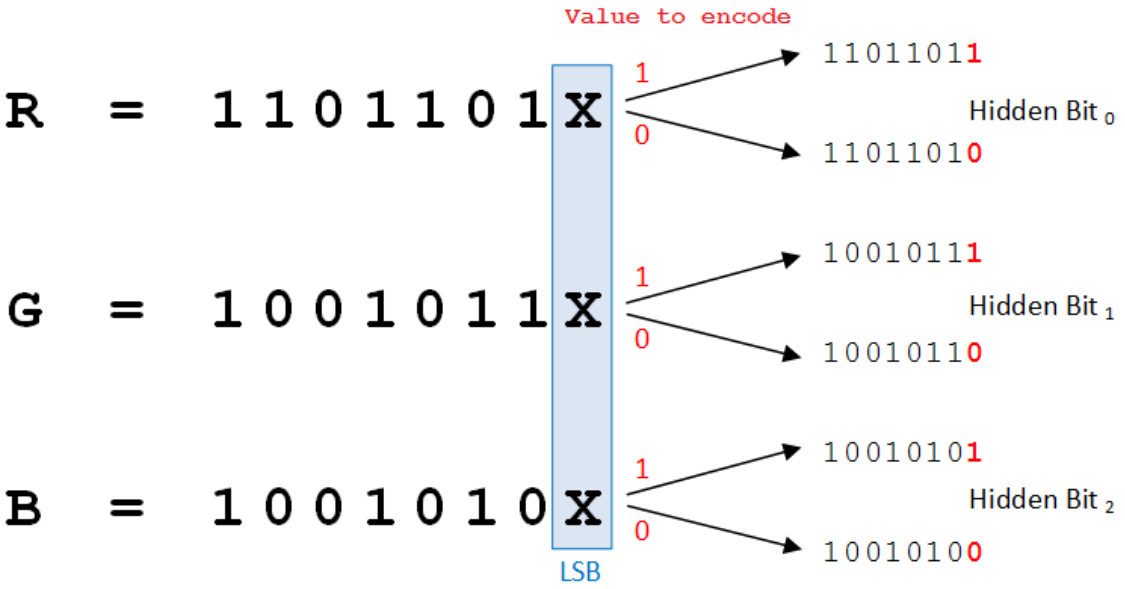
- **Stegoloader** uses the most popular (and the simplest) digital media steganography technique: **Least Significant Bit** (LSB) modification
- Stegoloader has a modular design and steganography is utilized to **hide its main module's code** inside a Portable Network Graphics (PNG) image downloaded from a legitimate website

LSB: A Common Steganographic Technique



RGB(218,150,149)

R = 11011010
G = 10010110
B = 10010101



Malware that injects secret data **into network traffic: W32/Foreign.LXES!tr**

- **W32/Foreign.LXES!tr** hides malicious traffic under the *HTTP/1.1 404 Not Found* Error

- **HTTP 404 Error** is a standard HTTP response code that indicates that the client is able to communicate to a server but that the **server could not find the page that the client is requesting**

W32/Foreign.LXES!tr

- Data exchange with C&C server is hidden in the **HTTP 404 Error** within the source code comment between the NCMD keywords

Recent examples of information hiding-based malware

Turla Backdoor Deployed in Attacks Against Worldwide Targets

By Sergiu Gatlan

The hacking group controls the backdoor using specially crafted JPG or PDF attachments which contain the commands encoded using steganography, with the backdoor executing its masters' orders and then automatically blocking emails containing commands upon detection.

Because the commands are encoded within attachments using steganography—the process of hiding information in plain sight by replacing bits that otherwise would be unused—even if the recipients would receive the 'control' emails.

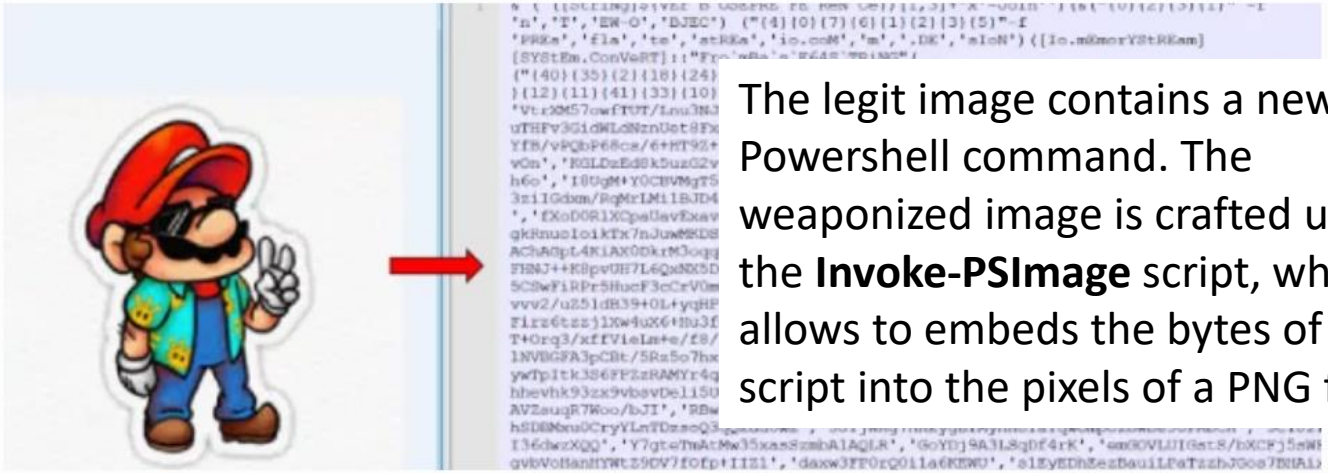
LSB JPG steganography

"In the case of a PDF, the command data can be anywhere in the document. LightNeuron operators just add a header at the beginning of the PDF to specify the offset at which the data is located," says ESET.

PDF steganography

Recent examples of information hiding-based malware

Ursnif: Long Live the Steganography!



The legit image contains a new Powershell command. The weaponized image is crafted using the **Invoke-PSImage** script, which allows to embeds the bytes of a script into the pixels of a PNG file.

Introduction

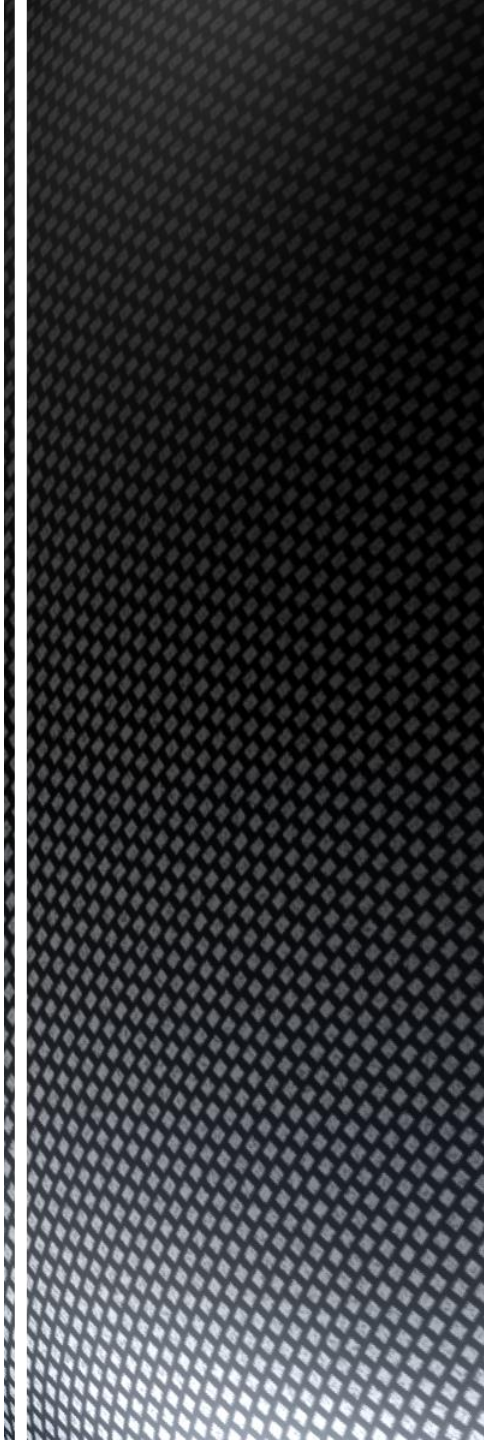
Another wave of Ursnif attacks hits Italy.

Ursnif is one of the most active banking trojan. It is also known as GOZI, in fact it is a fork of the original Gozi-ISFB banking Trojan that got its source code leaked in 2014 updating and evolving Gozi features over the years. Also in this variant, Ursnif use weaponized office document with a VBA macro embedded that act as a dropper and multi-stage highly obfuscated powershell scripts in order to hide the real payload. In addition, this Ursnif use also steganography to hide the malicious code and avoid AV detection.

Information hiding in future malware

- What we observe in current malware today **is not suprising** at all for experts & academics
- These information hiding techniques **have been known for years**
- And there are **many more** sophisticated methods available
- Application of the information hiding techniques will lead to even **more sophisticated and stealthier malware** that will be even harder to detect than nowadays
- First symptoms have already been observed – a new type of **advanced persistent threat** (APT) and **now ordinary malware follows**

Trends in network covert channels

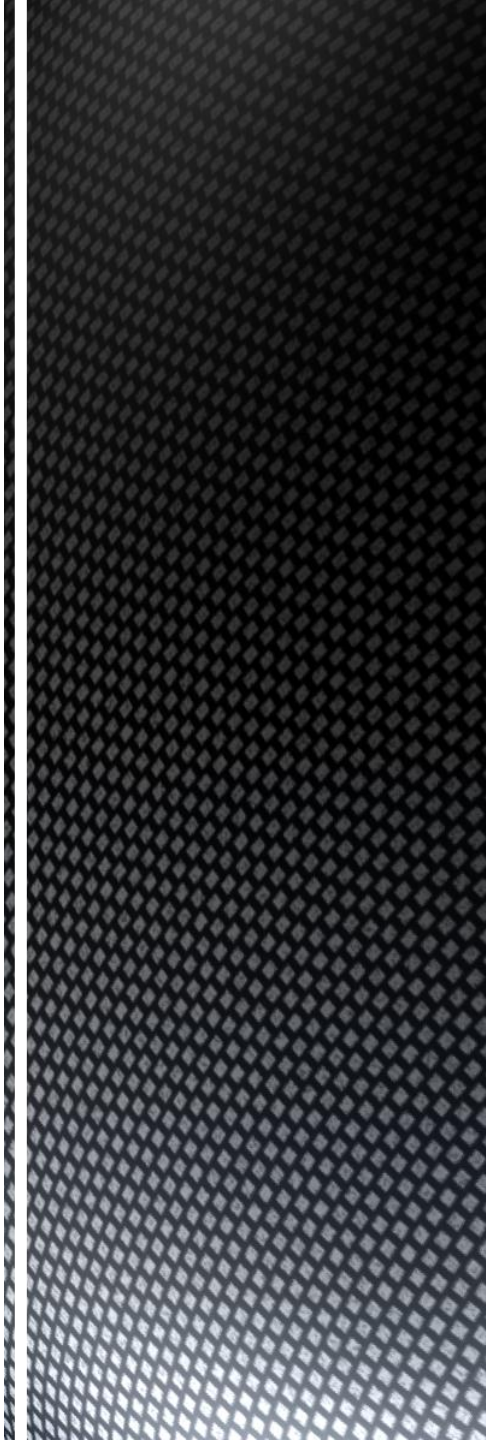


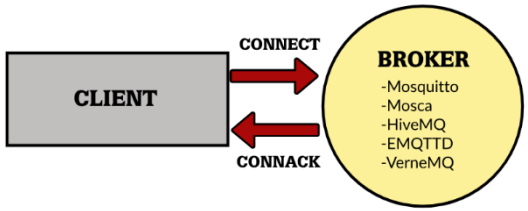
Information hiding in future malware: potential trends

Potential developments for information hiding-capable future malware:

- **Better digital media steganography algorithms** – they are already available: e.g. F5, HUGO – harder to detect/eliminate
- **New devices/networking environments:** smartphones, Internet of Things (IoT), CPS, SDN, etc.
- **New services/protocols:** Dropbox, Skype, VoIP, BitTorrent, SCTP etc.
- **New concept for botnets:** overlay network that utilizes only steganographic methods to communicate (stego-botnets)

New Networking Environments for Covert Channel: **IoT (Internet of Things)**





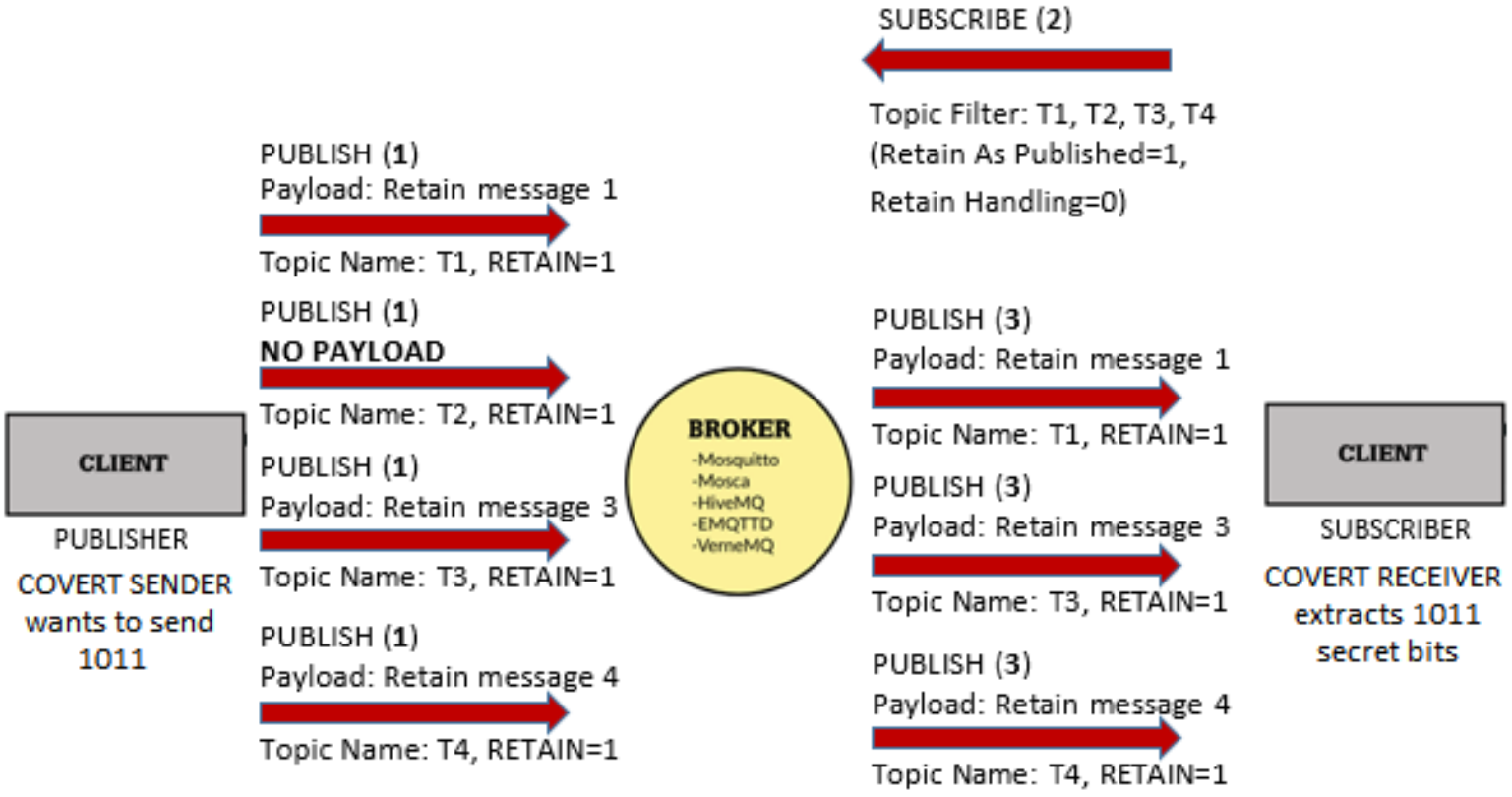
IoT: MQTT-based Covert Channels

- **Message Queuing Telemetry Transport (MQTT)** is currently widely deployed in Internet of Things (IoT) environments
- A first comprehensive study of covert channels in MQTT (a **publish-subscriber model**-based protocol)
- We identified in total **13 covert channels**: 7 direct and 6 indirect covert channels
- We prove MQTT-based covert practical feasibility by implementing the chosen data hiding scheme and perform its **experimental evaluation**

*A. Velinov, A. Mileva, S. Wendzel, W. Mazurczyk - **Covert Channels in MQTT-based Internet of Things**, IEEE Access, 2019, DOI: 10.1109/ACCESS.2019.2951425*

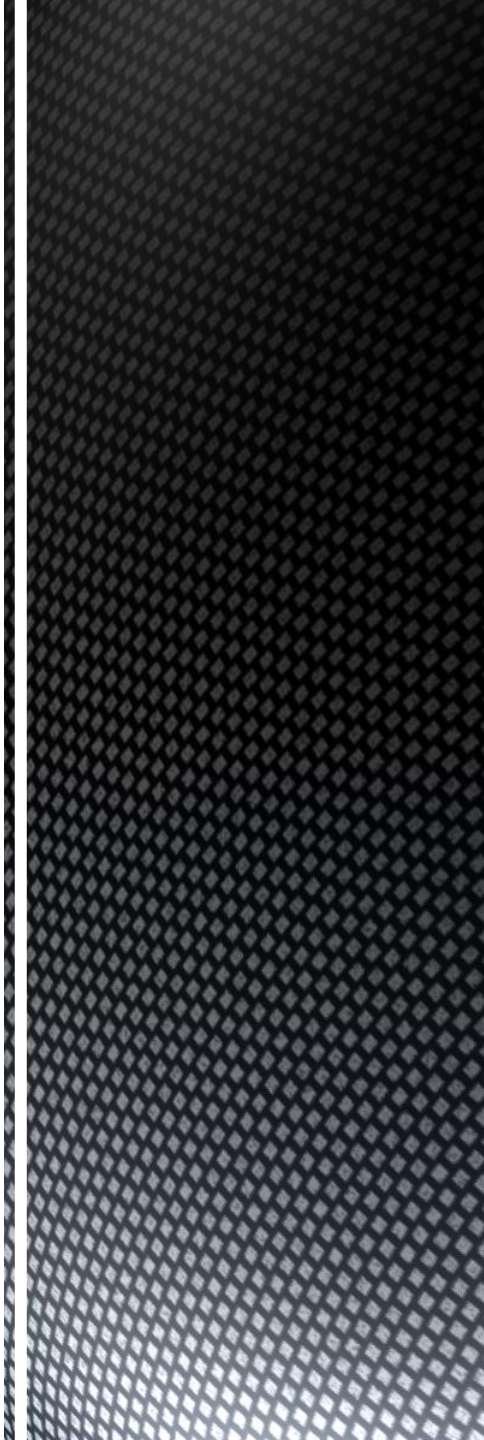
IoT: MQTT-based Covert Channels

Indirect Covert Channel using Topic Ordering and Updates Presence/Absence

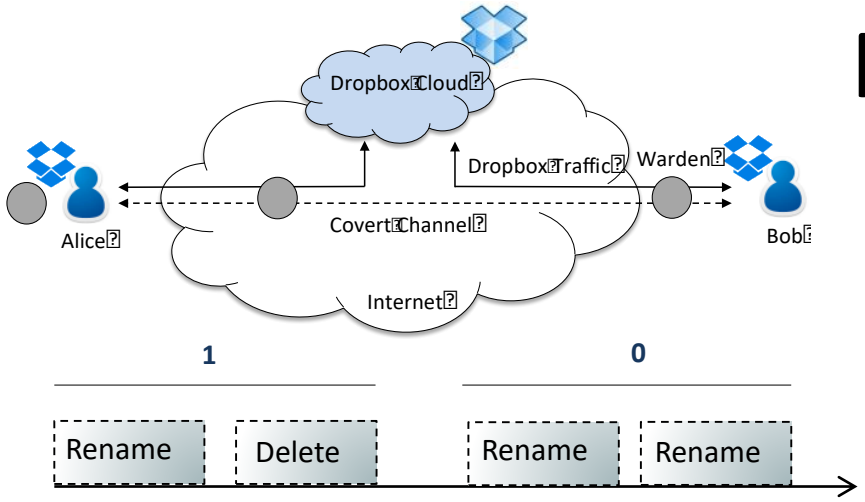


*A. Velinov, A. Mileva, S. Wendzel, W. Mazurczyk - **Covert Channels in MQTT-based Internet of Things**, IEEE Access, 2019, DOI: 10.1109/ACCESS.2019.2951425*

Network Covert Channels in **Personal Cloud Storage**

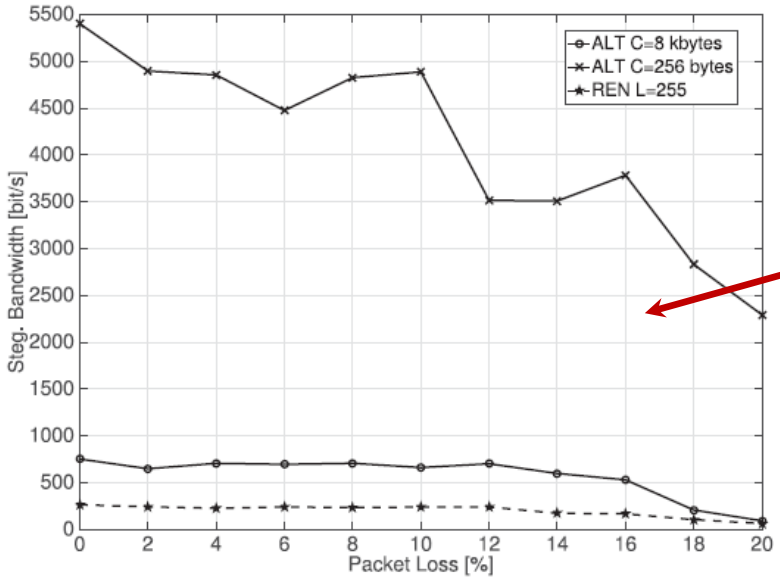


Personal Cloud Storage



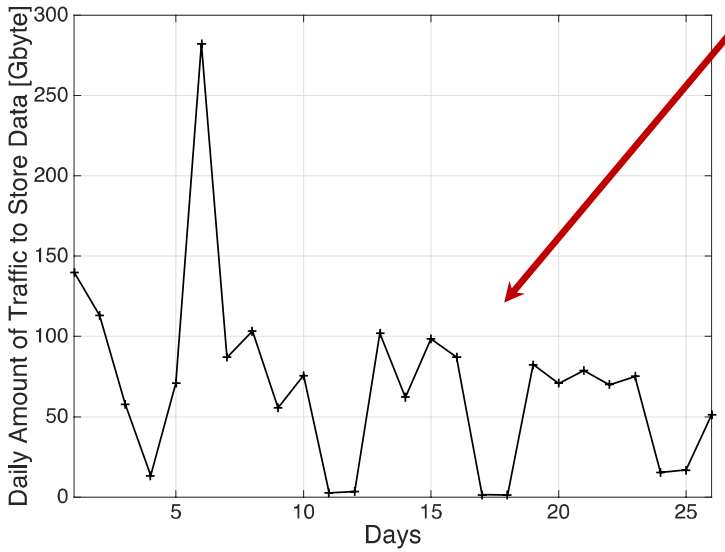
- **Cloud architectures** or services can be also used to create covert channels
- An example uses Dropbox – key idea:
  - Changing the content of a file will produce a “a file has changed” notification;
  - Changing the name of a file will produce a “file name has changed” notification;
  - Add/delete file(s) will produce a “Y file(s) has (have) been added/removed”;
  - This can be used to encode bits into patterns of operations.
- We implemented two methods using Dropbox and evaluated their performances **in realistic scenarios**

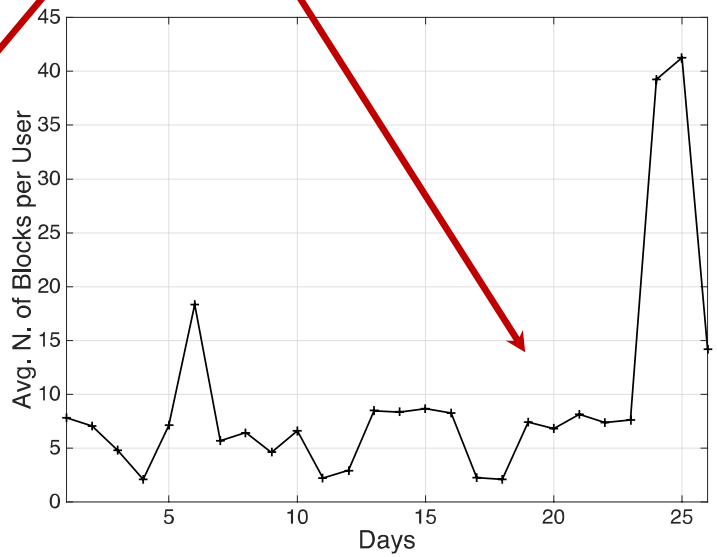
Personal Cloud Storage



The resulting bandwidth depends on the method and how aggressively the method is used

In order not to be easily spotted the data hiding must take into account users' behavior



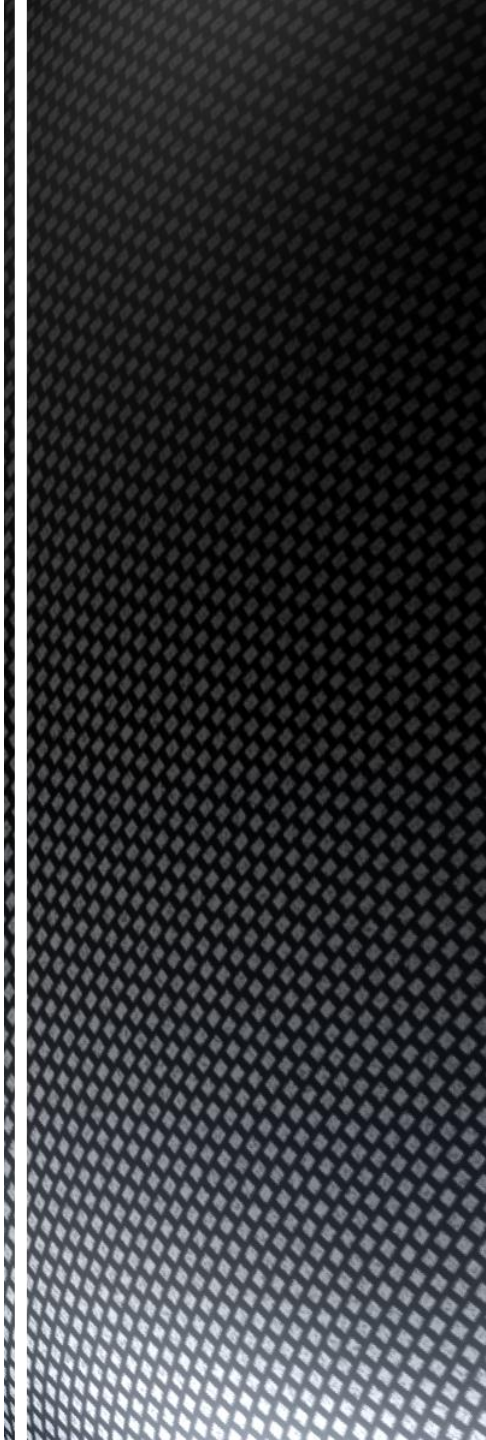


The most important part of this research is not the development of a new covert channel. Instead, creating new methods allows to increase the understanding of what can be exploited in a complex scenario, as well as particular user behaviors to develop effective countermeasures and detection techniques.

L. Caviglione, M. Podolski, W. Mazurczyk, M. Ianigro - ***Covert Channels in Personal Cloud Storage Services: the case of Dropbox****, IEEE Transactions on Industrial Informatics, Vol. 13, Iss. 4, pp. 1921-1931, 2017 DOI: 10.1109/TII.2016.2627503*

VoIP Network Covert Channel: TranSteg

(Transcoding Steganography)



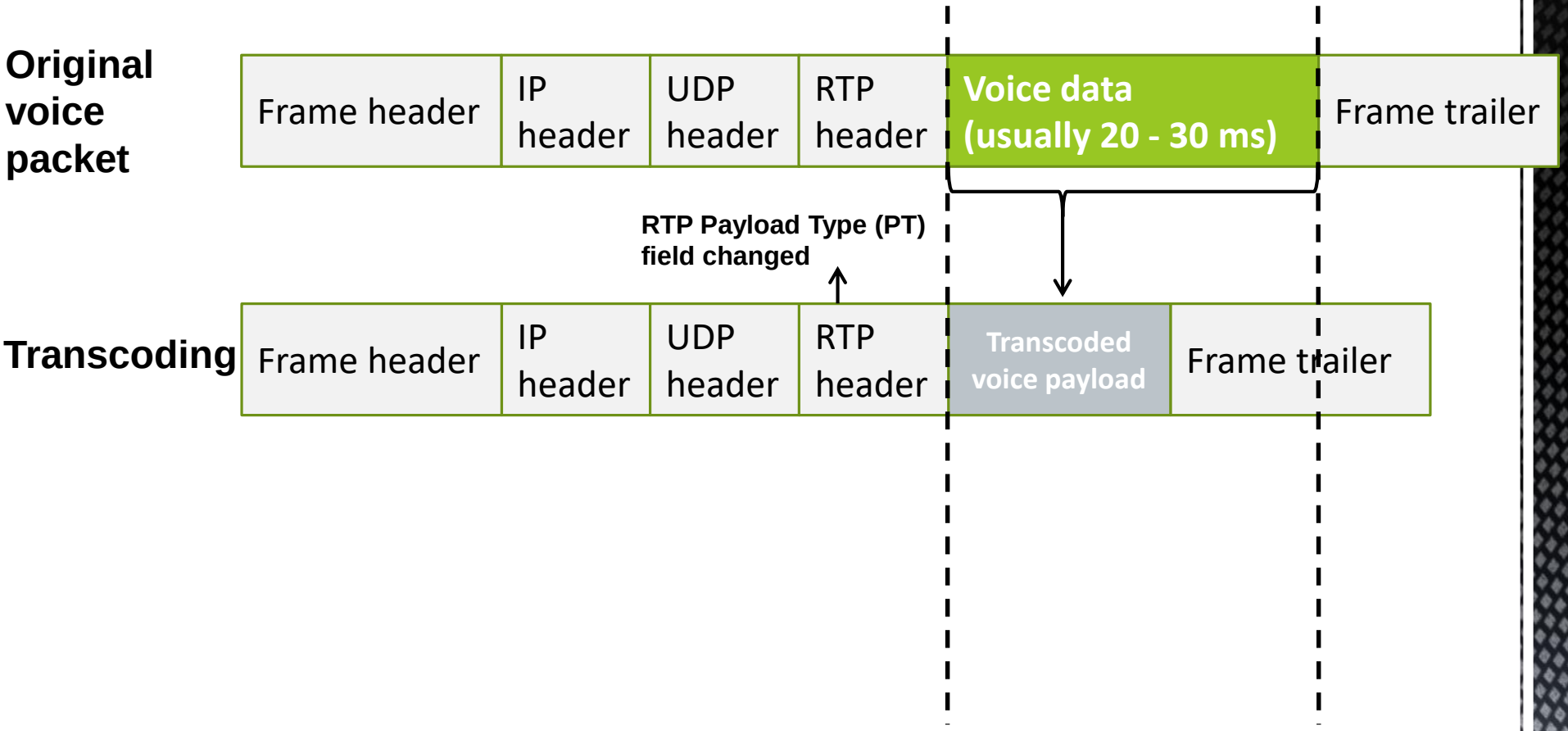
TranSteg features

- TranSteg is intended for a broad class of multimedia and real-time applications e.g. **IP telephony** or services like video streaming
- The typical approach to steganography is **to compress the covert data in order to limit its size** because it is reasonable in the context of a limited steganographic bandwidth
- TranSteg utilizes **compression of the overt data to make space for the secret data**
- TranSteg for IP Telephony is using transcoding of the voice data from a higher bit rate codec – ***overt codec*** to a lower bit rate codec – ***covert codec*** with the least possible degradation in voice quality

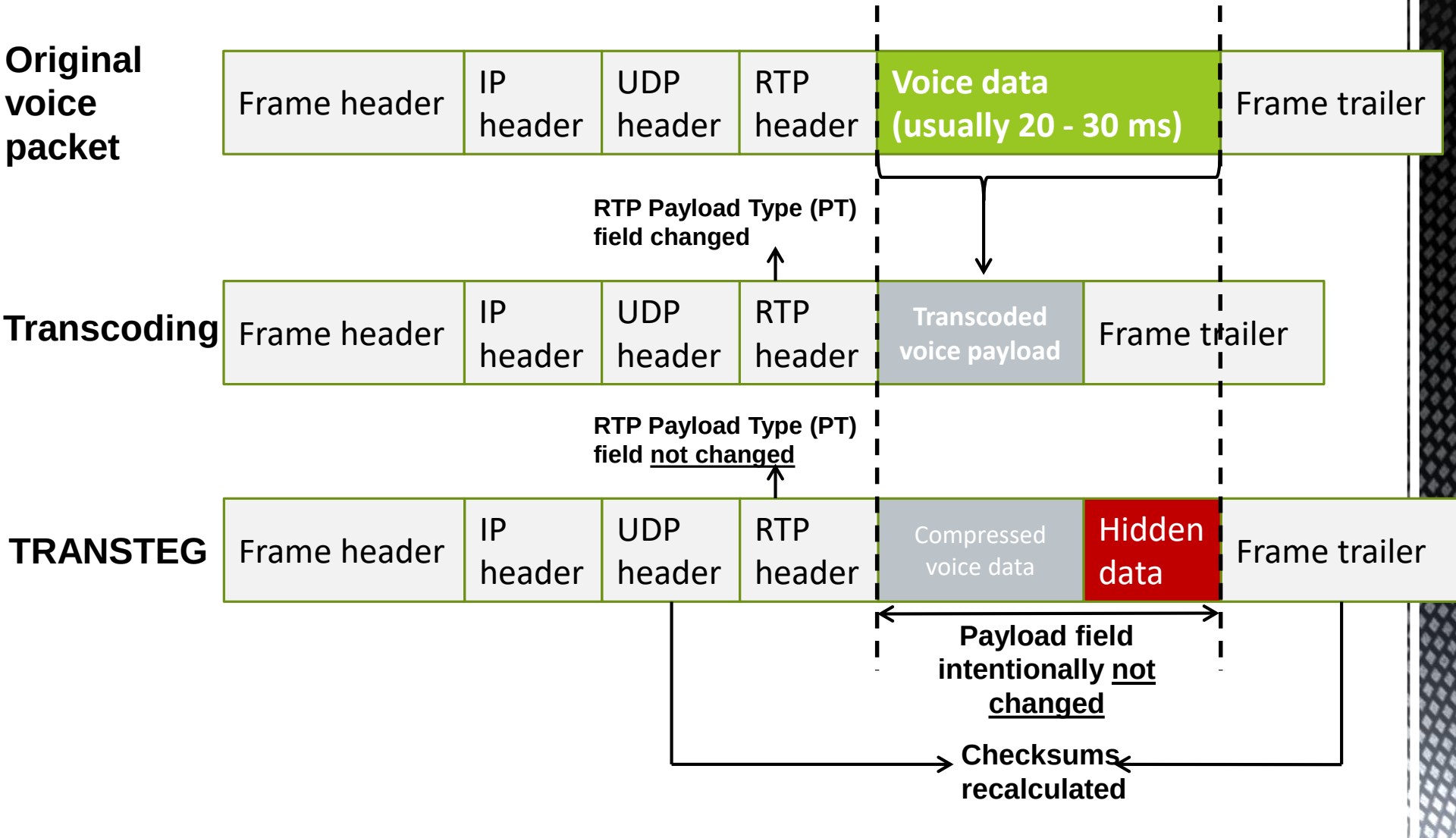
TranSteg in action (1/2)

Original voice packet

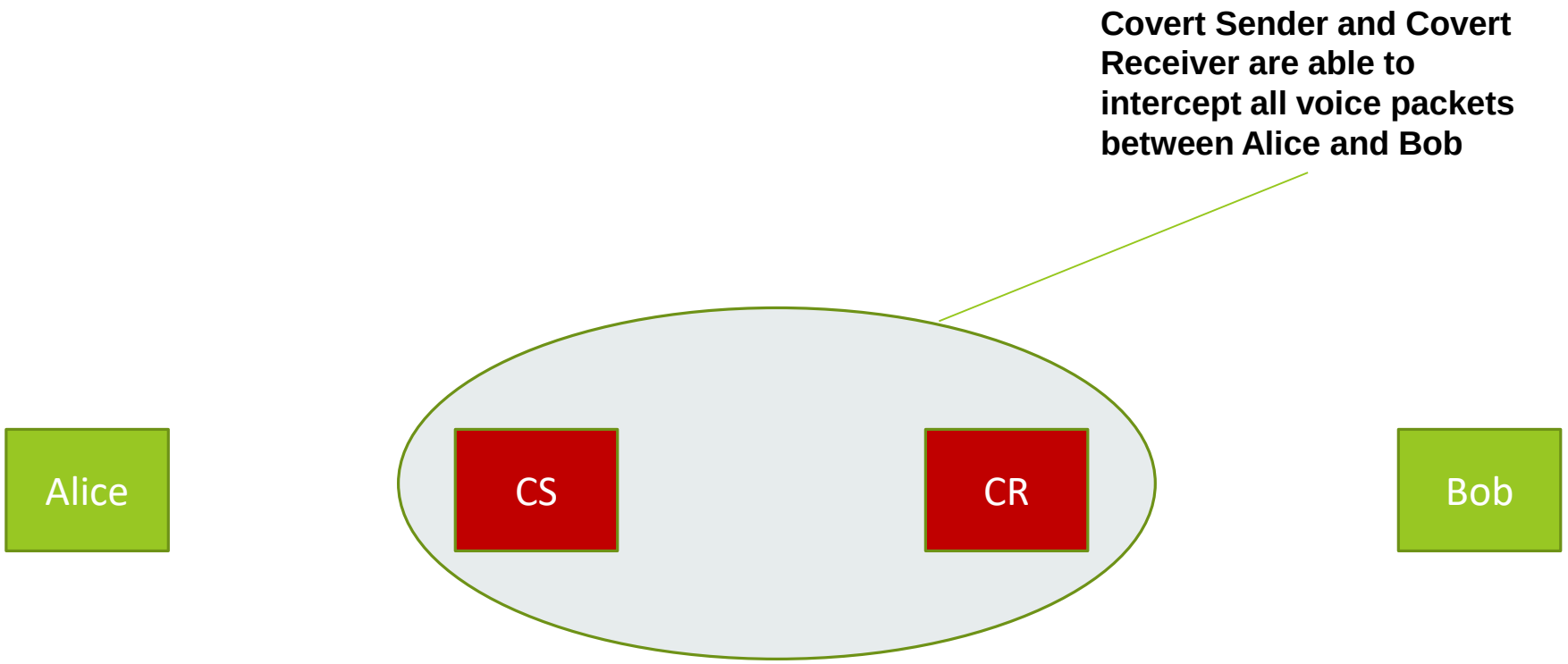
TranSteg in action (1/2)



TranSteg in action (1/2)



TranSteg in action (2/2)



CS – Covert Sender

CR – Covert Receiver

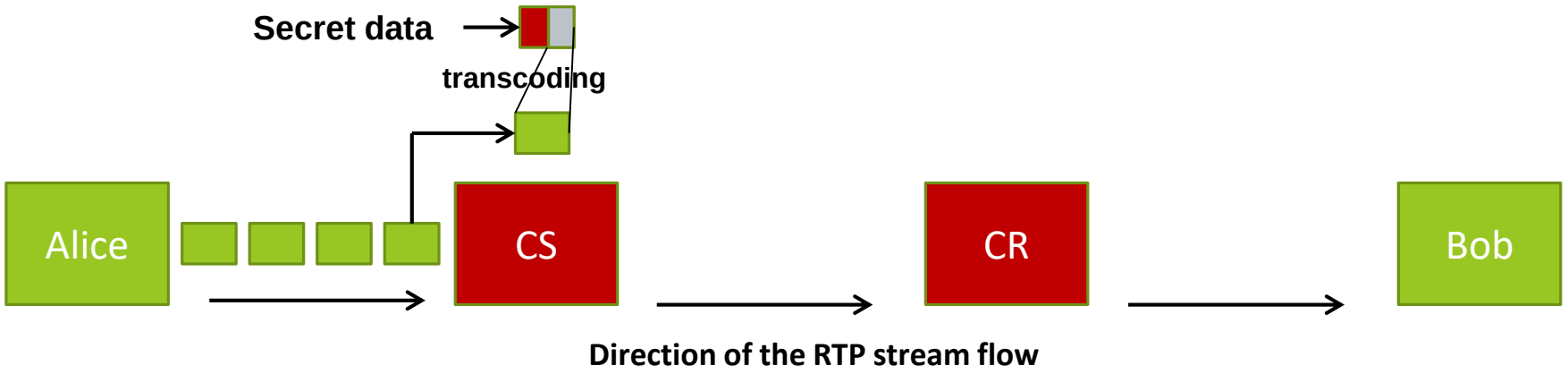
TranSteg in action (2/2)

Direction of the RTP stream flow

CS – Covert Sender

CR – Covert Receiver

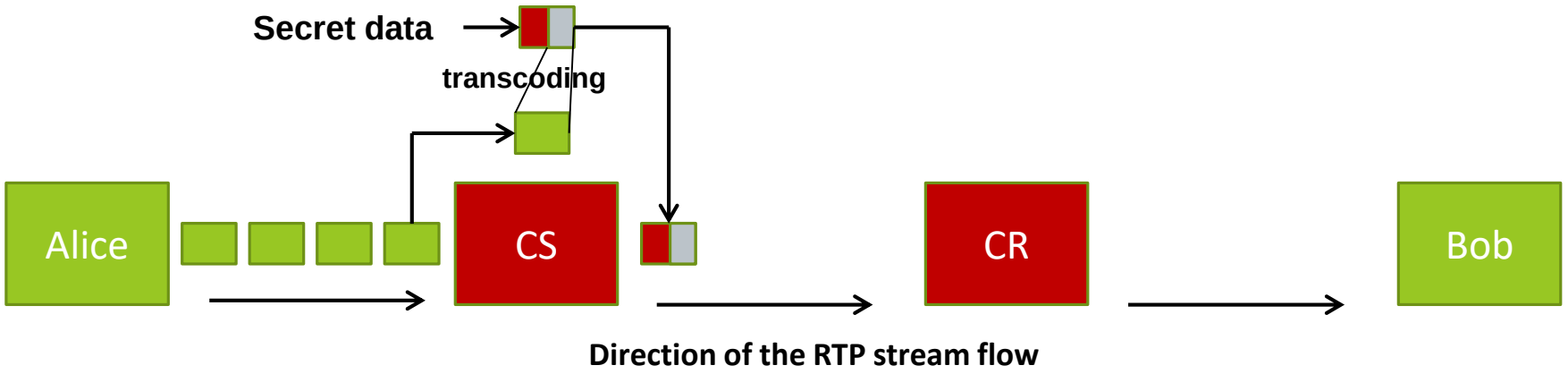
TranSteg in action (2/2)



CS – Covert Sender

CR – Covert Receiver

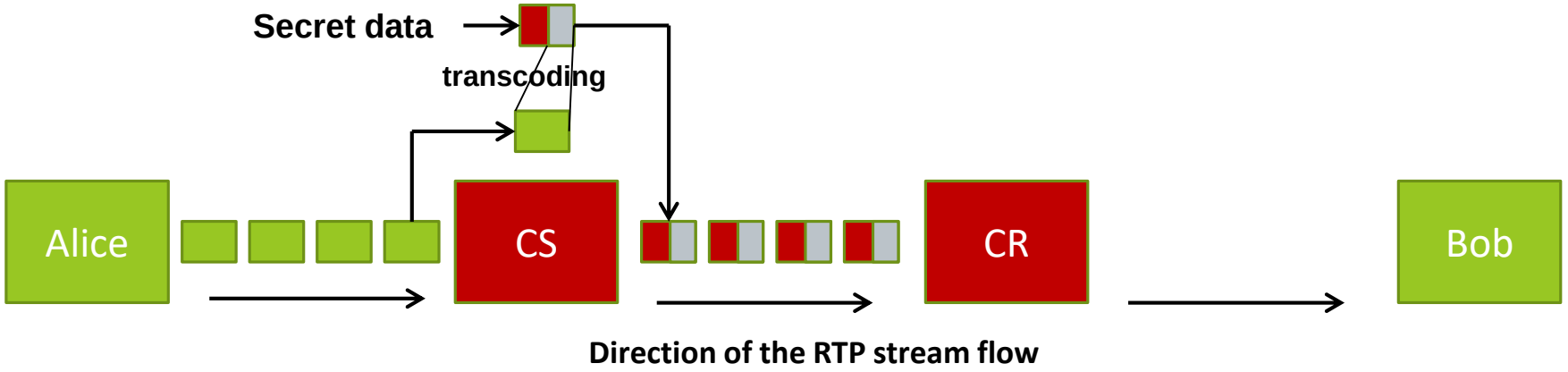
TranSteg in action (2/2)



CS – Covert Sender

CR – Covert Receiver

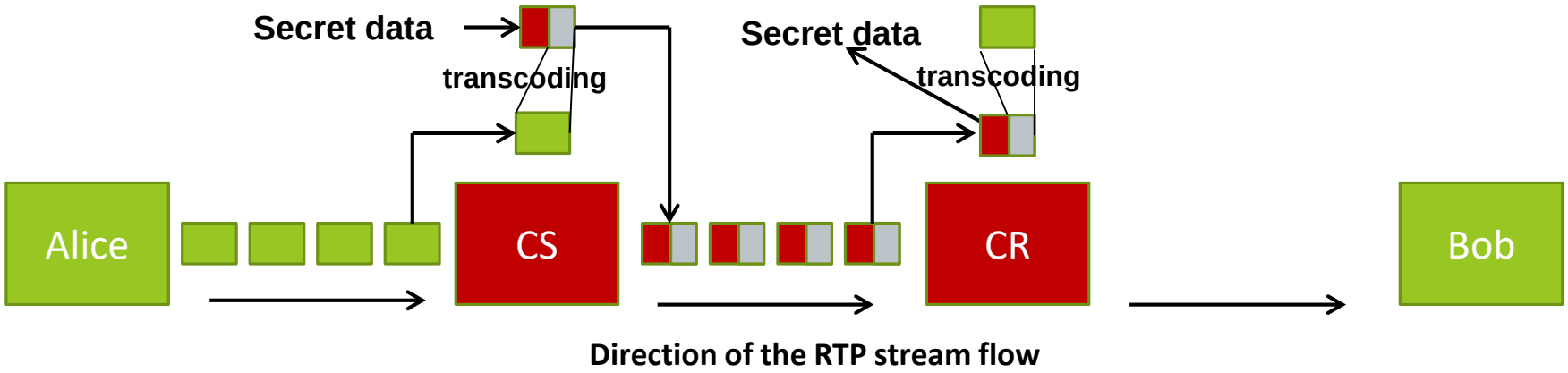
TranSteg in action (2/2)



CS – Covert Sender

CR – Covert Receiver

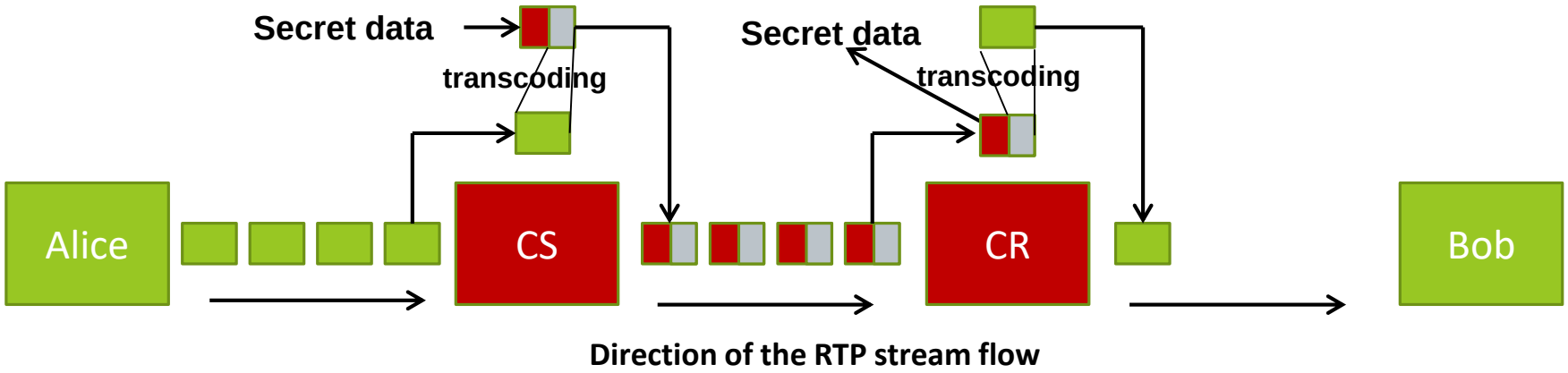
TranSteg in action (2/2)



CS – Covert Sender

CR – Covert Receiver

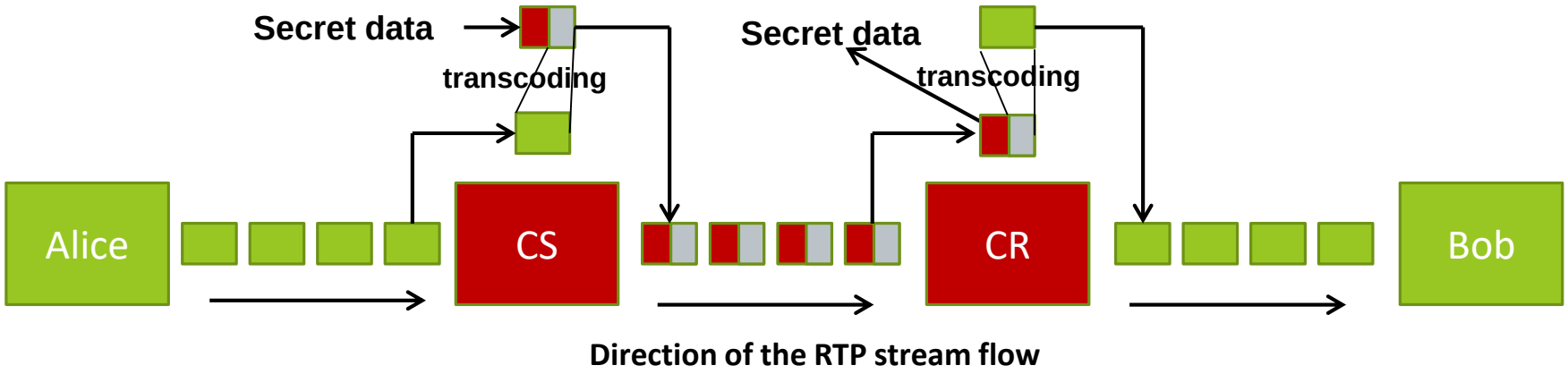
TranSteg in action (2/2)



CS – Covert Sender

CR – Covert Receiver

TranSteg in action (2/2)



CS – Covert Sender

CR – Covert Receiver

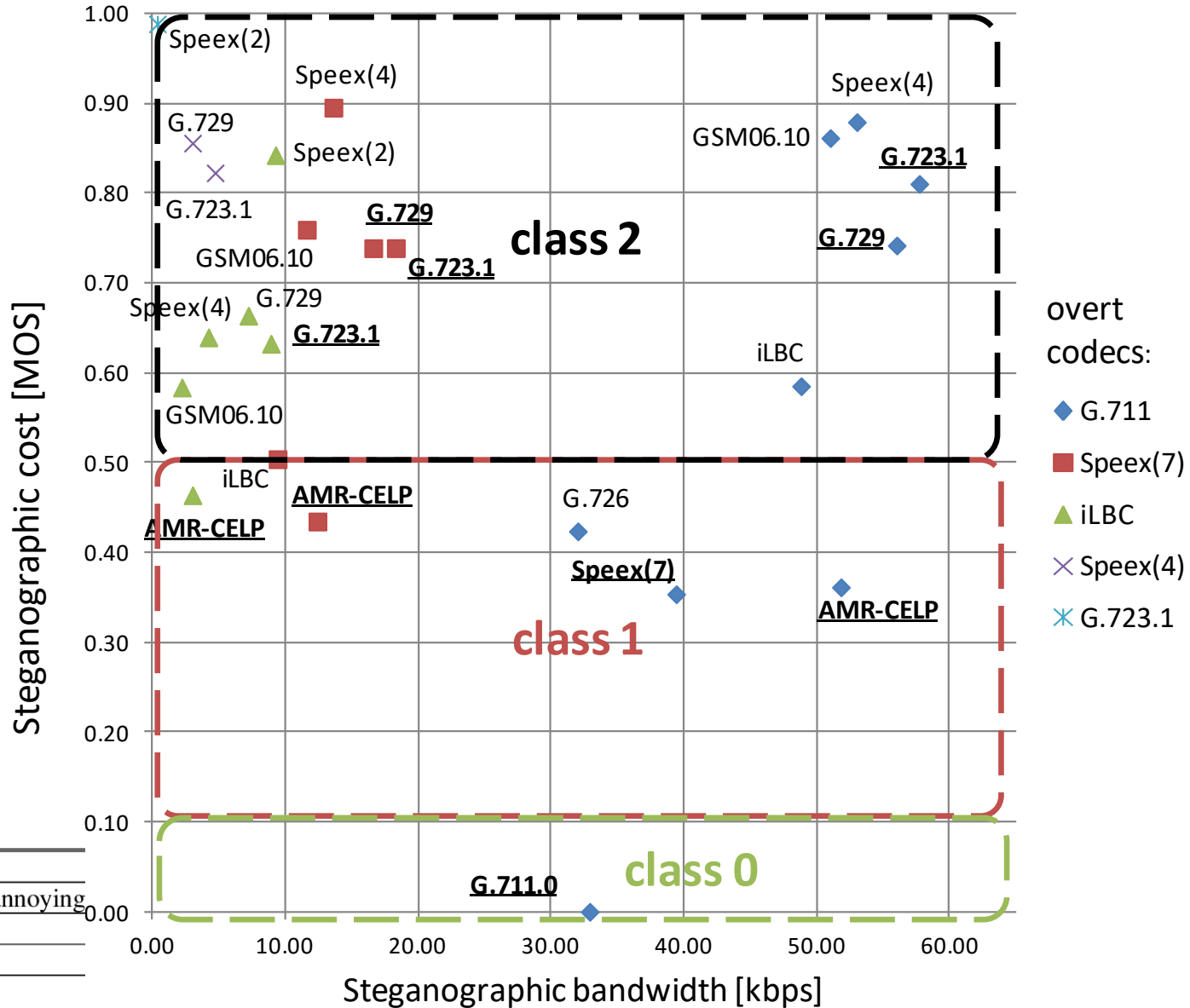
Initial TranSteg Results

- TranSteg experimental results for a pair of codecs: **G.711** (overt codec: 64kbit/s) and **G.726** (covert codec: 32 kbit/s)
- A high steganographic bandwidth – **32 kbit/s** – was achieved while introducing delays lower than **1 ms**, and still retaining good voice quality (**14,4 MB/hour; 345 MB/day; 10.4 GB/month**)
- Detection strongly depends on the realized **hidden communication scenario** and the capabilities of a warden responsible for network steganography detection
- Generally TranSteg detection is difficult to perform especially if the **voice stream is encrypted**

*W. Mazurczyk, P. Szaga, K. Szczypiorski - **Using Transcoding for Hidden Communication in IP Telephony** - Multimedia Tools and Applications, Volume 70, Issue 3, pp. 2139-2165, DOI: 10.1007/s11042-012-1224-8, June 2014*

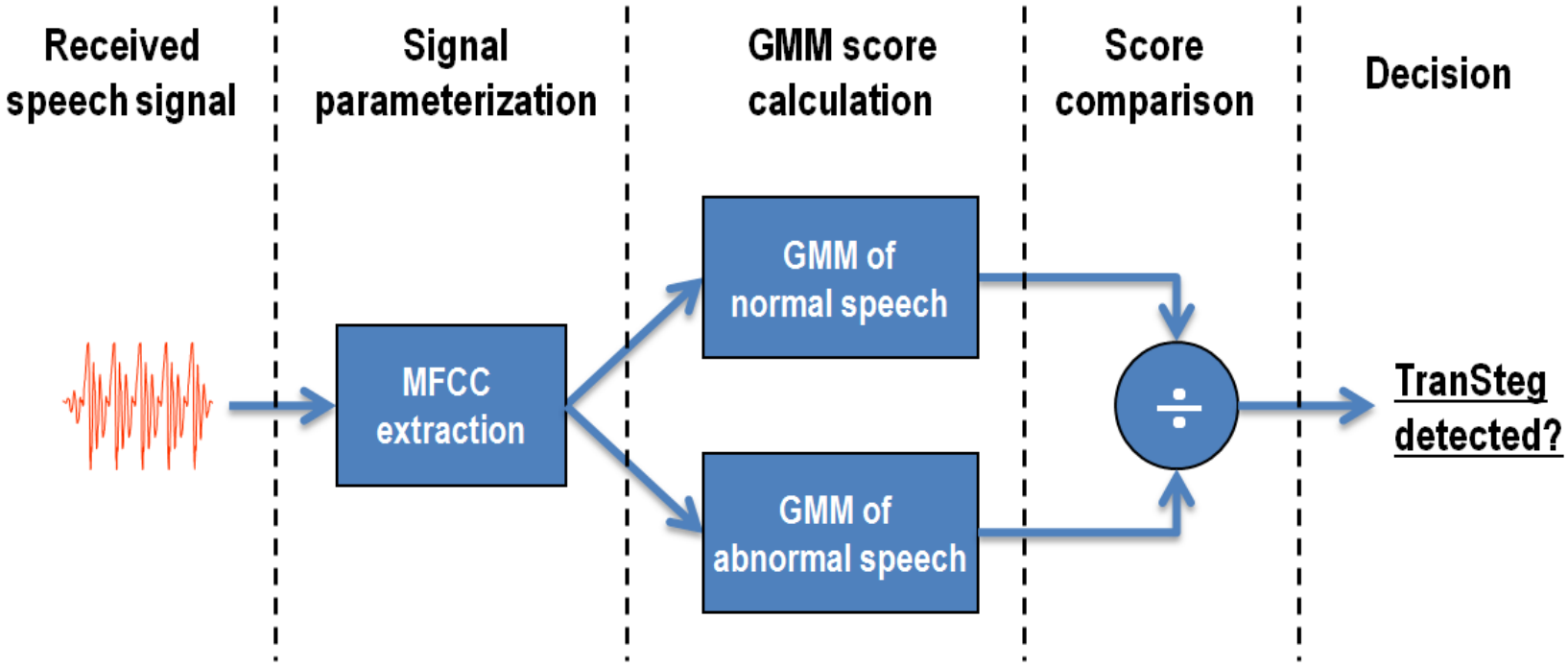
*A. Janicki, W. Mazurczyk, K. Szczypiorski - **Influence of Speech Codecs Selection on Transcoding Steganography** - Telecommunication Systems: Modelling, Analysis, Design and Management, Vol. 59, Iss. 3, pp. 305-3315, 2015*

TranSteg results



MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

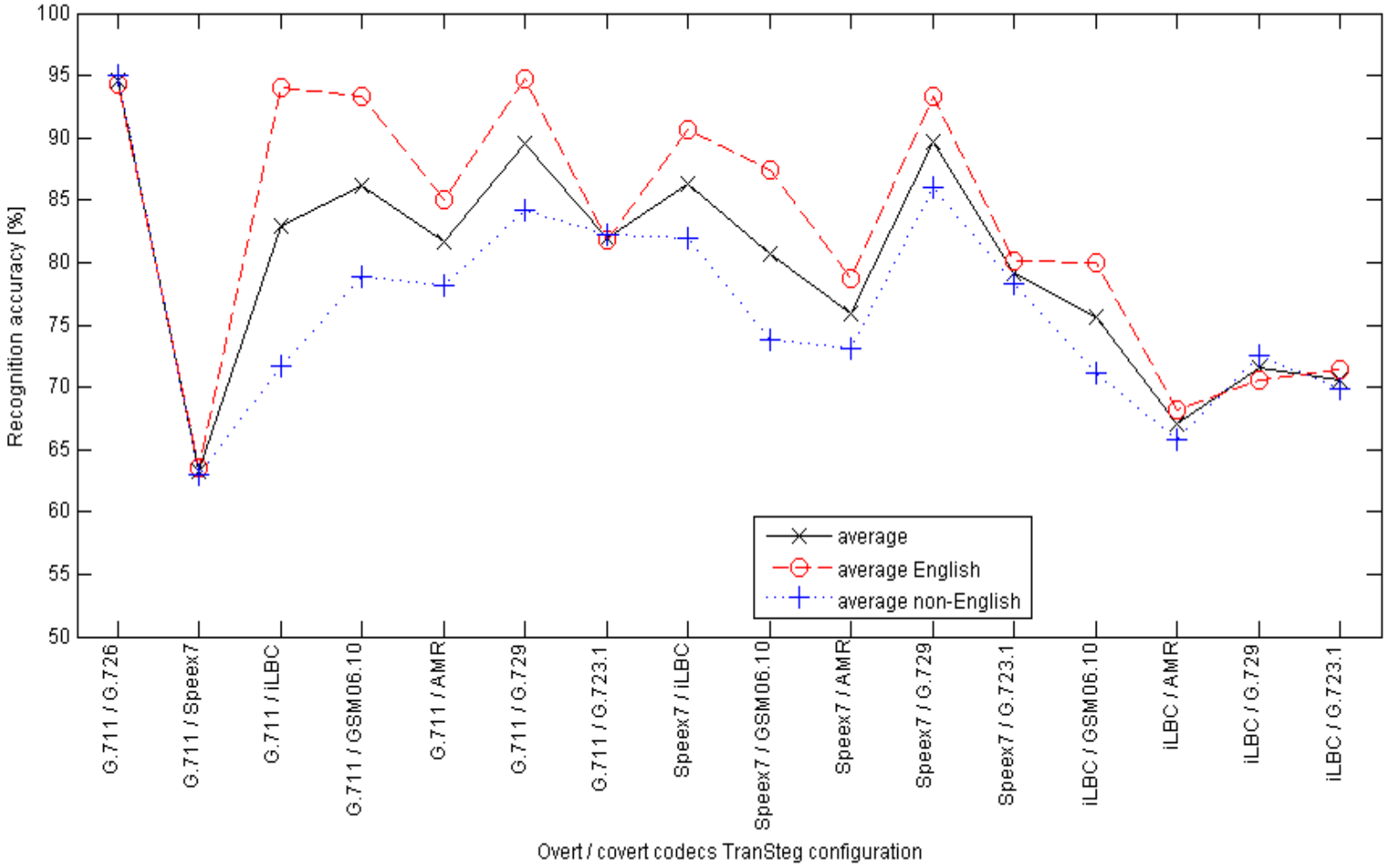
TranSteg detection



MFCC = Mel-Frequency Cepstral Coefficients
GMM = Gaussian Mixture Models

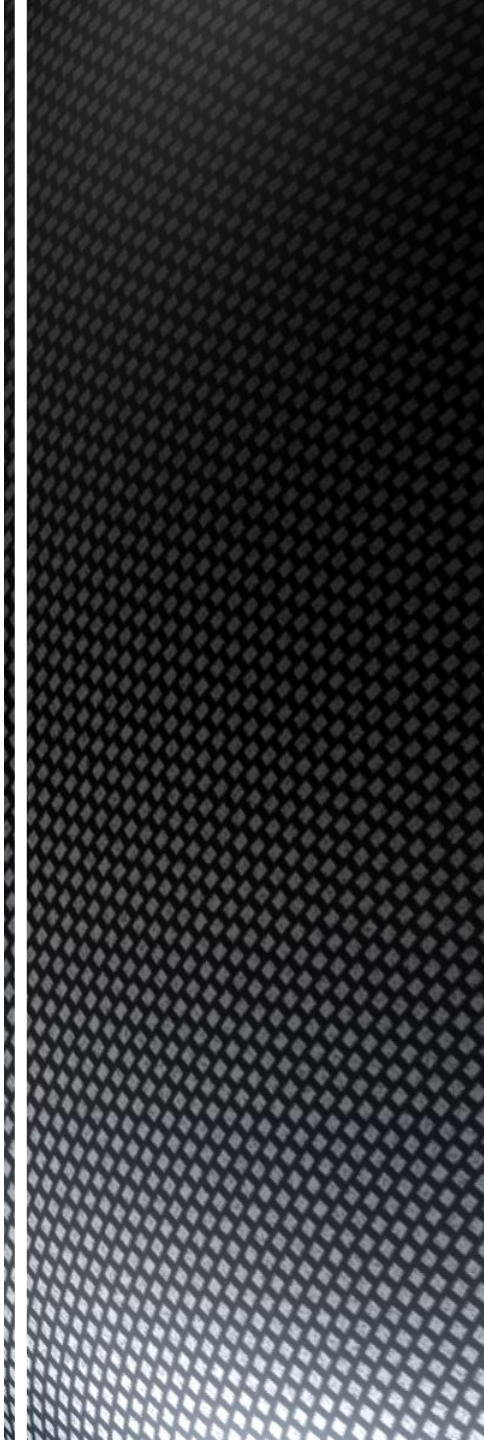
A. Janicki, W. Mazurczyk, K. Szczypiorski - Steganalysis of Transcoding Steganography - Annals of Telecommunications, Vol. 69, Iss. 7, 2014, pp. 449-460, DOI: 10.1007/s12243-013-0385-4

TranSteg recognition accuracy



A. Janicki, W. Mazurczyk, K. Szczypiorski - Steganalysis of Transcoding Steganography - Annals of Telecommunications, Vol. 69, Iss. 7, 2014, pp. 449-460, DOI: 10.1007/s12243-013-0385-4

Challenges for countering network information hiding



Challenges in network covert channels detection

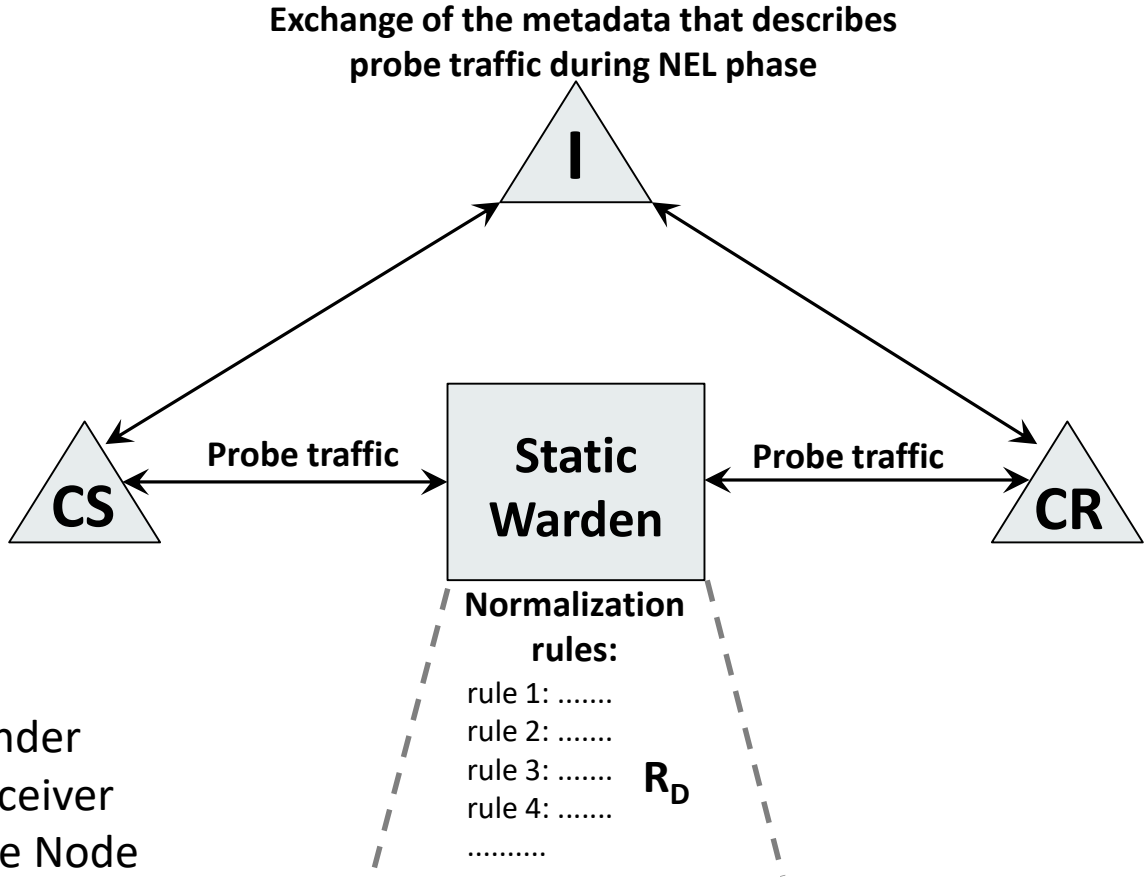
- **Asymmetry**: ease of development of new techniques vs. a challenge to detect/eliminate/limit
- Developing **effective and more general tool** to detect information hiding techniques in communication networks is **still an open challenge**
- Many of the existing detection methods **are not practically feasible** in current communication networks
- Currently (more or less):

  one steganographic method ≈ one detection solution

- It is **easier to (blindly) prevent** information hiding technique utilization than **to detect covert communication**

Regular (static) warden issue

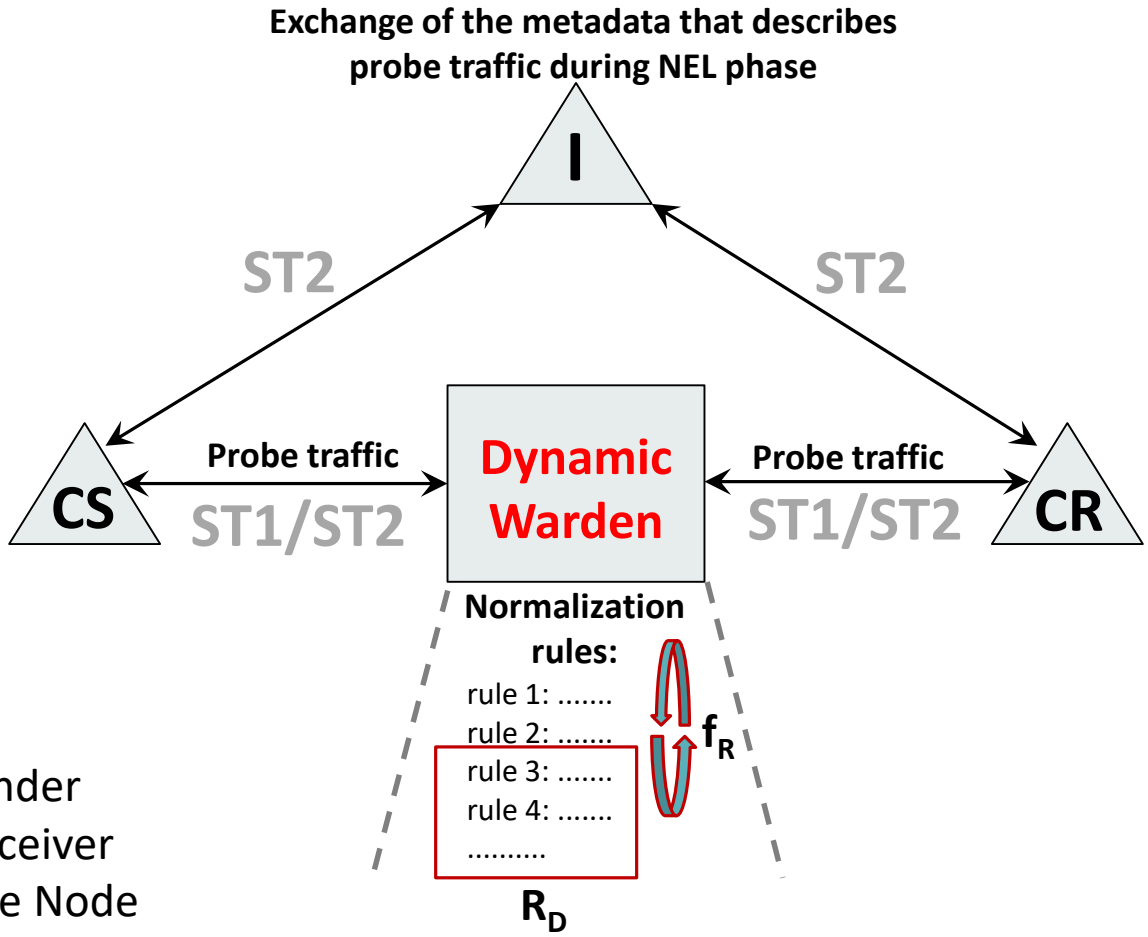
- ***Adaptive covert communication parties*** – improved information hiding-based threat - two stages: the network environment learning (**NEL**) phase and covert communication (**COM**) phase
- CS and CR are able to **infer normalization rules** used by the regular warden and then adapt/choose the data hiding technique **not covered by the warden**



CS – Covert Sender
CR – Covert Receiver
I – Intermediate Node

Dynamic Warden concept (Moving Target Defense)

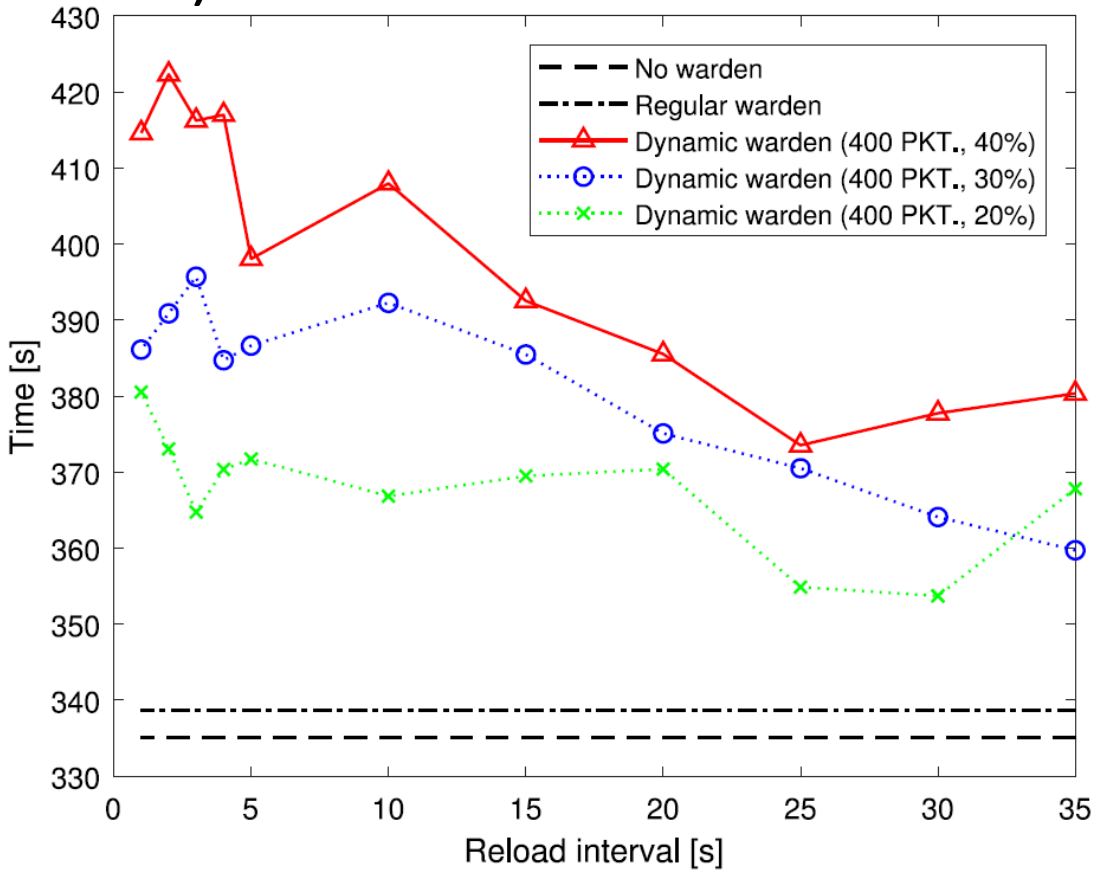
- New approach to elimination of network covert channels
- Focused to **deter adaptive covert communication** scenario by constantly shuffling the active normalization rules set **to confuse covert communication parties so the hidden data exchange to last longer**



CS – Covert Sender
CR – Covert Receiver
I – Intermediate Node

Dynamic Warden concept

- Experimental results prove that Dynamic Warden strategy makes attack to last **25% longer** (more time to spot covert communication)



*W. Mazurczyk, S. Wendzel, M. Chourib, J. Keller - **Countering Network Covert Channels with Dynamic Wardens**, Future Generation Computer Systems, vol. 94, 712–725, 2019*

Future research directions

- Which complex and **advanced „constructions”** of network covert channels can be utilized by cyber criminals in the future?
- What makes some protocols/services **more prone to data hiding than others** (and how to express this)? How and why this susceptibility **changes in time**?
- Can we construct a protocol/service in such a way that it is **immune to information hiding** or ensure that this susceptibility is significantly limited?
- What should we take into account and how to design detection/prevention methods **so they are more universal/general**?

FernUniversität in Hagen

CUING

Trends in Stegomalware: Techniques and Countermeasures

Wojciech Mazurczyk, Ph.D., D.Sc.

FernUniversität in Hagen, Germany

3rd International Conference on Frontiers in Cyber Security (FCS 2020)

SIMARGL

18 December 2020 **Tianjin, P.R. China**