Secrecy, Stealth, Privacy and Storage for Noisy Channels and Identifiers

Gerhard Kramer
Technical University of Munich

Talk at the European School of Information Theory
Chalmers University, Gothenburg, Sweden
April 7, 2016

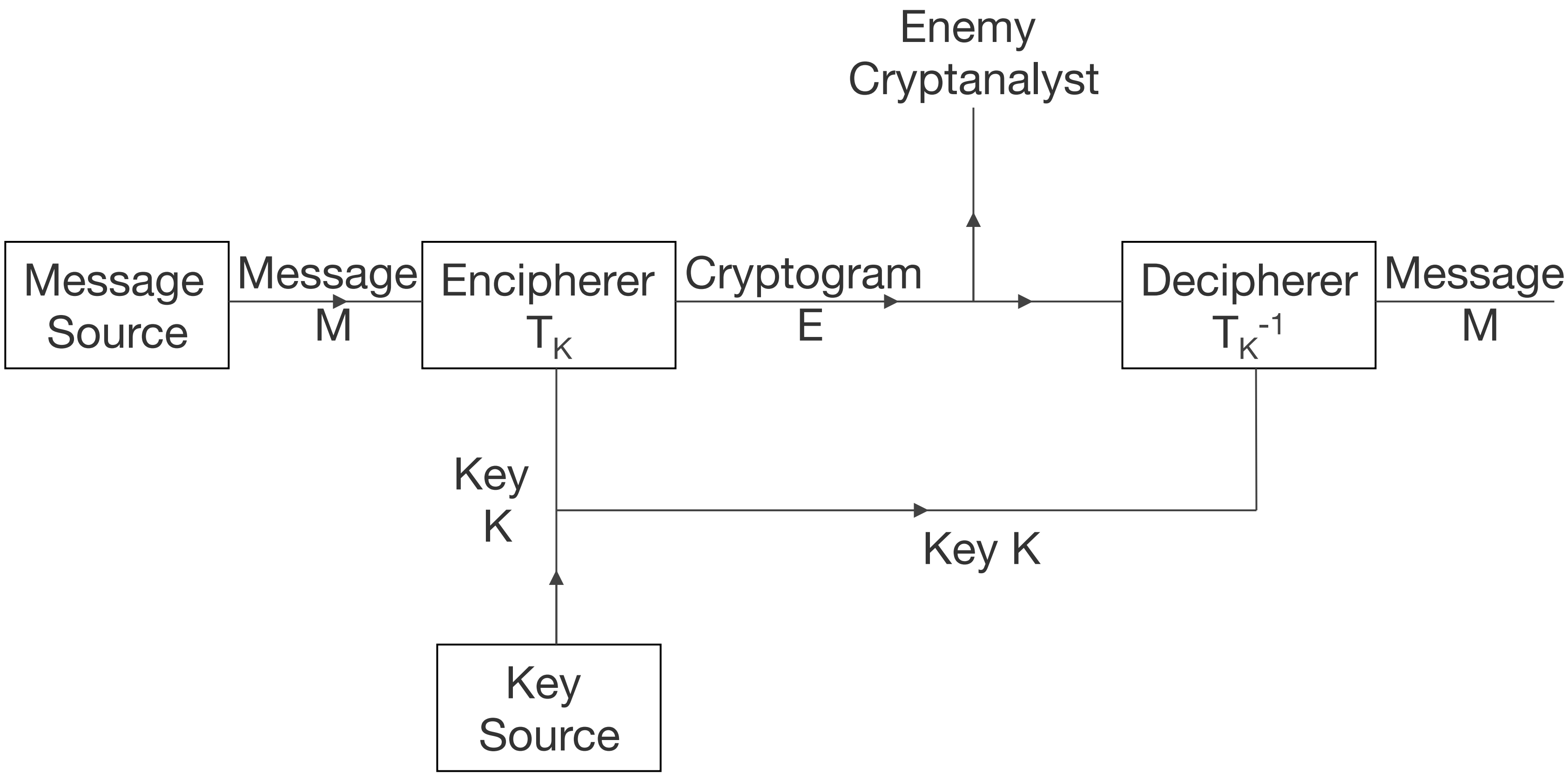
CHALMERS
UNIVERSITY OF TECHNOLOGY

Unterstützt von / Supported by
Alexander von Humboldt
Stiftung / Foundation

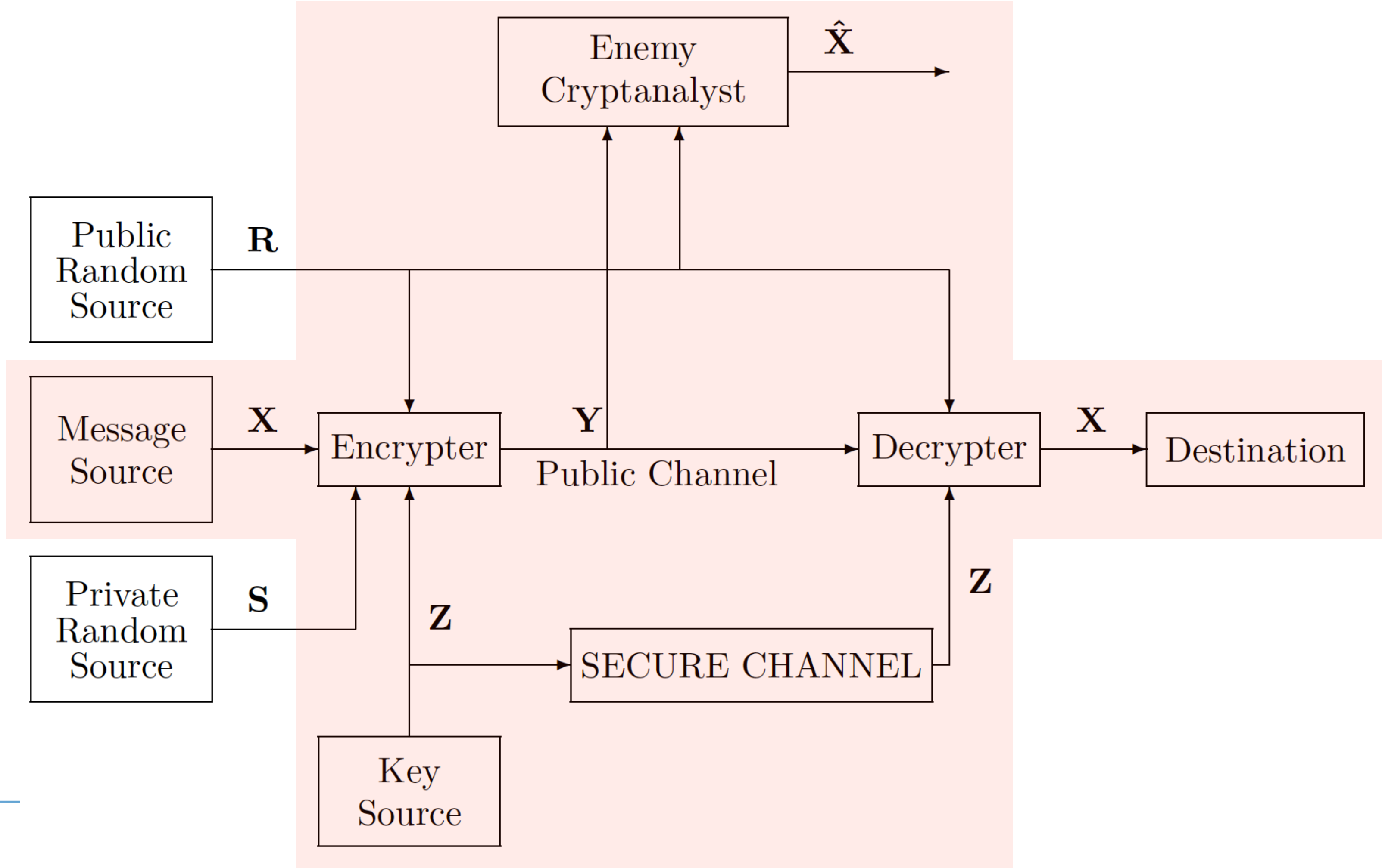
Motivation

Motivation: Secrecy

- Example 1: Shannon's schematic of a general secrecy system (Communication Theory of Secrecy Systems, BLTJ, 1949)



- Example 2: Massey's general model of a secrecy-key cryptosystem (ADIT 2 – ETH Course Notes 1981-97)



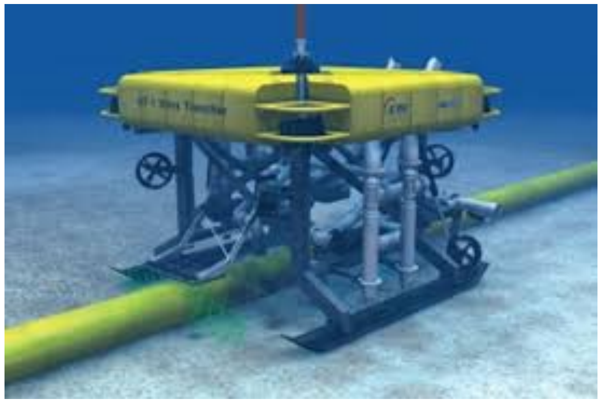
Motivation: Secrecy Without a Key

- Example 3: Wyner's Wiretap Channel

Source

Private Data

Wiretap



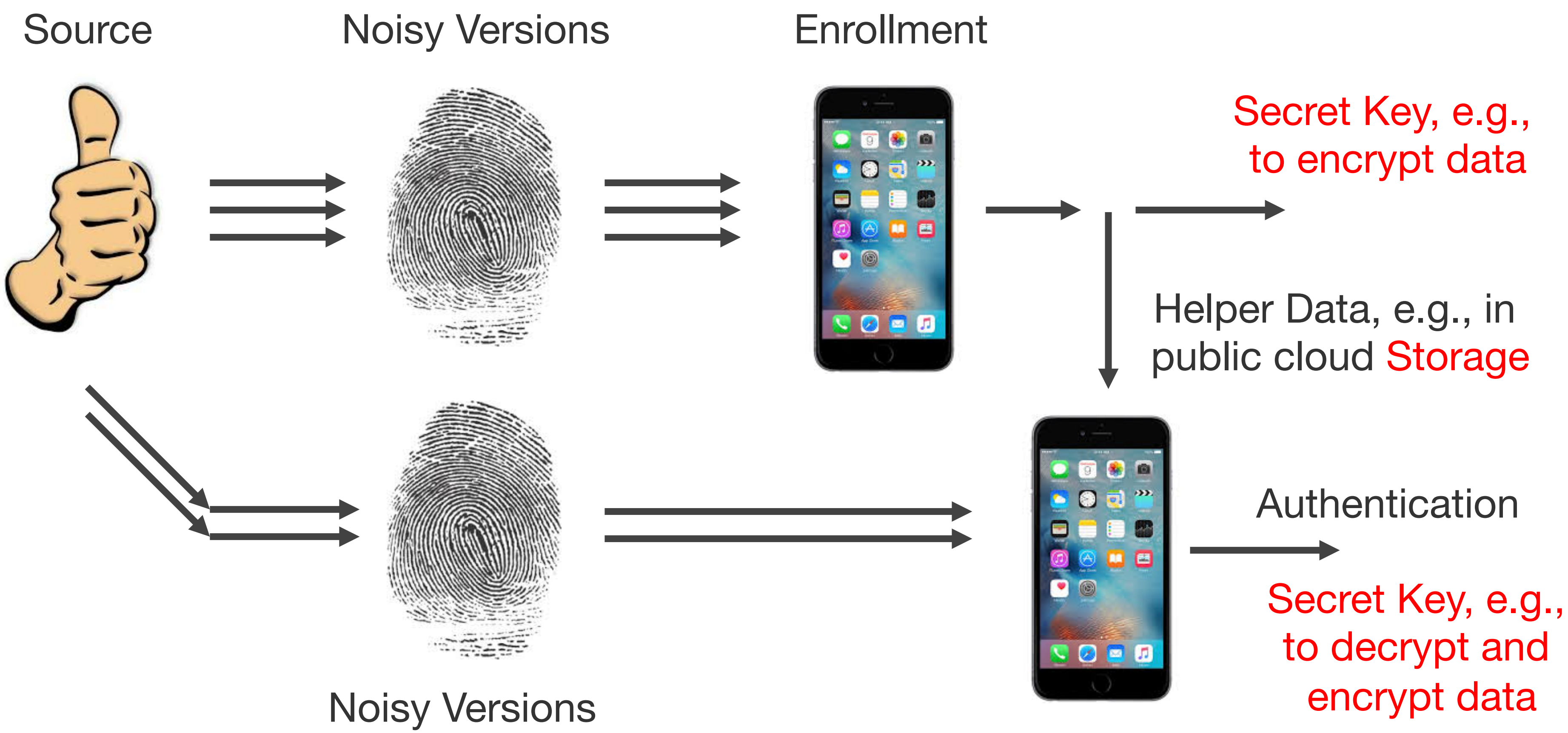
Private Data

Destination

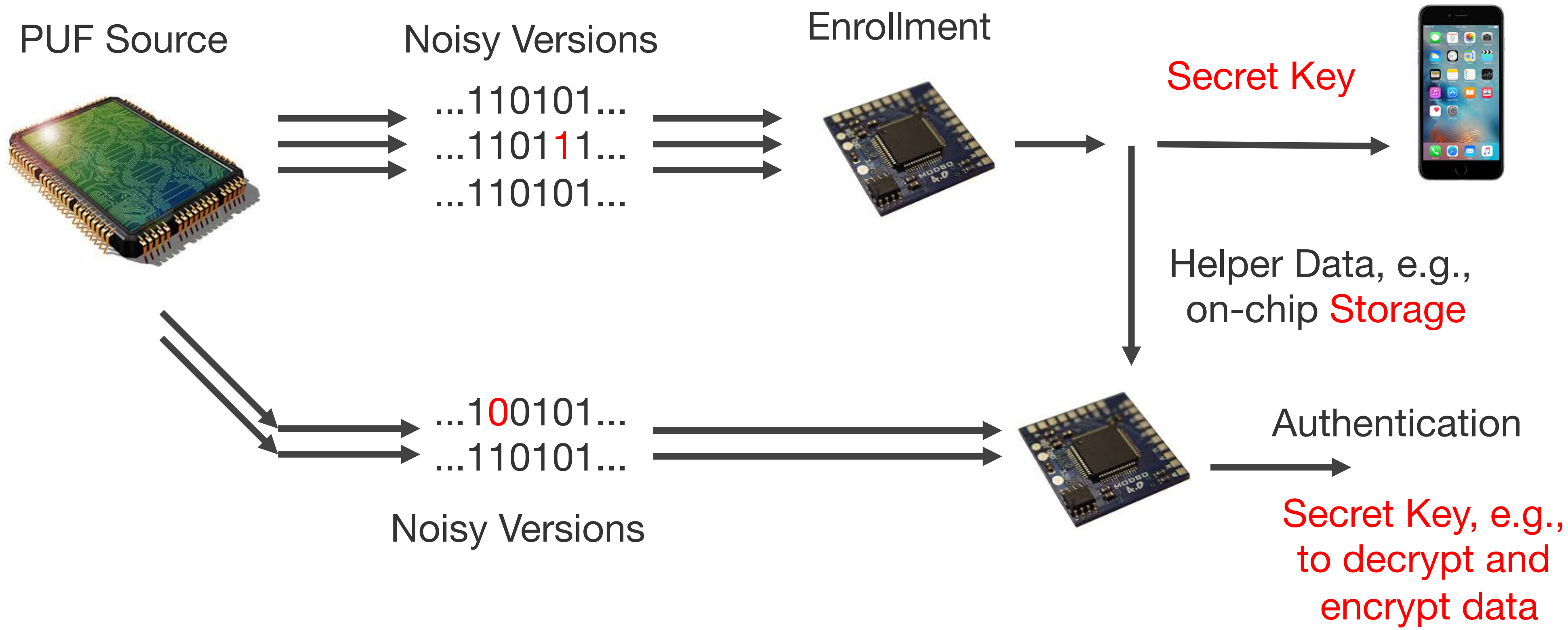
Wiretapper
wants the Data

Motivation: Secrecy and Privacy

- Example 4: Biometric Security



- Example 5: Device Security for Things and their Internet, Hardware “Fingerprint” via a PUF*



* Physical Unclonable Function

Motivation: Stealth

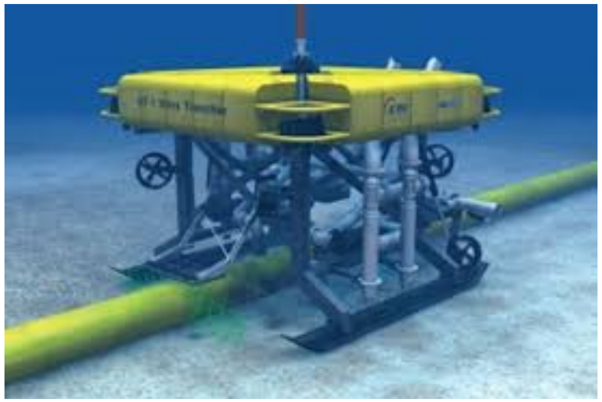
- Example 6: Wiretap Channel with a New Requirement

Source

Public Data →

Private Data

Wiretap



Public Data →

Private Data

Destination

↓

Wiretapper
wants to know <u>if</u> Private Data
was sent and, if yes, decrypt it

Stealth (Discussion)

- Low Probability of Intercept (LPI):* communication methods whose primary purpose is to prevent an unauthorized listener from determining the presence or location of the transmitter, in order to decrease the possibility of both electronic attack (jamming) and physical attack

* Based on Prescott 1993 (AFSOR Grant #AFOSR-91-0018)

Stealth (Discussion Continued)

- Four sequential operations that exploitation systems attempt to perform:

  1) Cover: a receiver is tuned to frequencies occupied by a signal of interest
  2) Detect: decide whether the signal is data plus noise and interference or just noise and interference.
  3) Intercept: extract features of the signal to determine if it is interesting
  4) Exploit: extract signal features as necessary and demodulate the baseband signal to generate a stream of (meaningful) binary digits.

- Interpretation: 4) deals with secrecy and 2) and/or 3) deal with stealth
- Example of 2): covert communication where data signal has very low energy
  Example of 3): some data signals may be uninteresting (see above)

Part 1:
Secrecy and Stealth for Wiretap Channels

Information Theory and “Basic” Models

Information Theory

- Entropy:

$$H(X) = \sum_{a \in \mathrm{supp}(P_X)} -P_X(a)\log P_X(a) = \mathrm{E}\left[-\log P_X(X)\right]$$

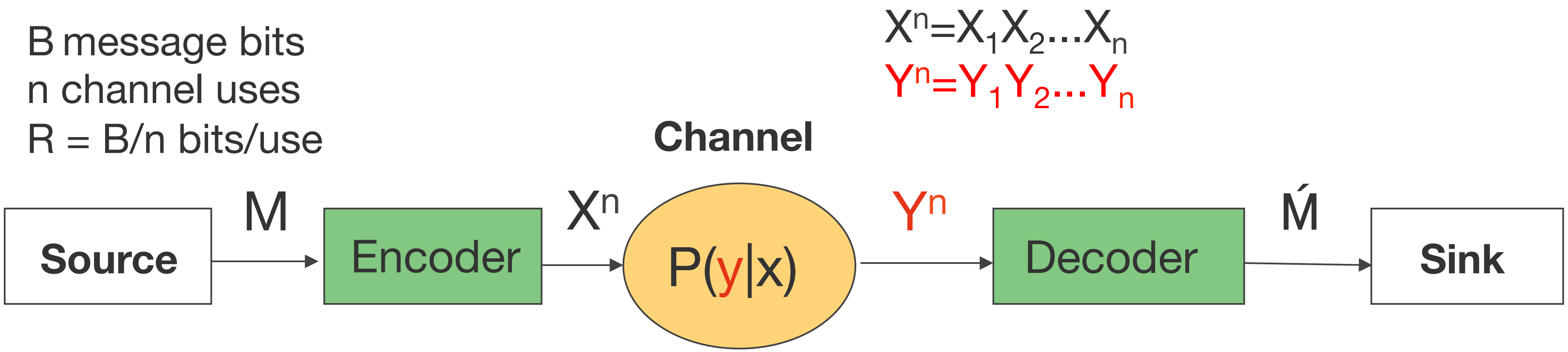
$$H(X|Y) = \sum_{ab \in \mathrm{supp}(P_{XY})} -P_{XY}(ab)\log P_{X|Y}(a|b)$$

- Mutual Information and Informational Divergence:

$$I(X;Y) = D(P_{XY} \| P_X P_Y)$$

$$= \sum_{ab \in \mathrm{supp}(P_{XY})} P_{XY}(ab)\log \frac{P_{XY}(ab)}{P_X(a)P_Y(b)} = \mathrm{E}\left[\log \frac{P_{XY}(XY)}{P_X(X)P_Y(Y)}\right]$$

Shannon's Channel Coding



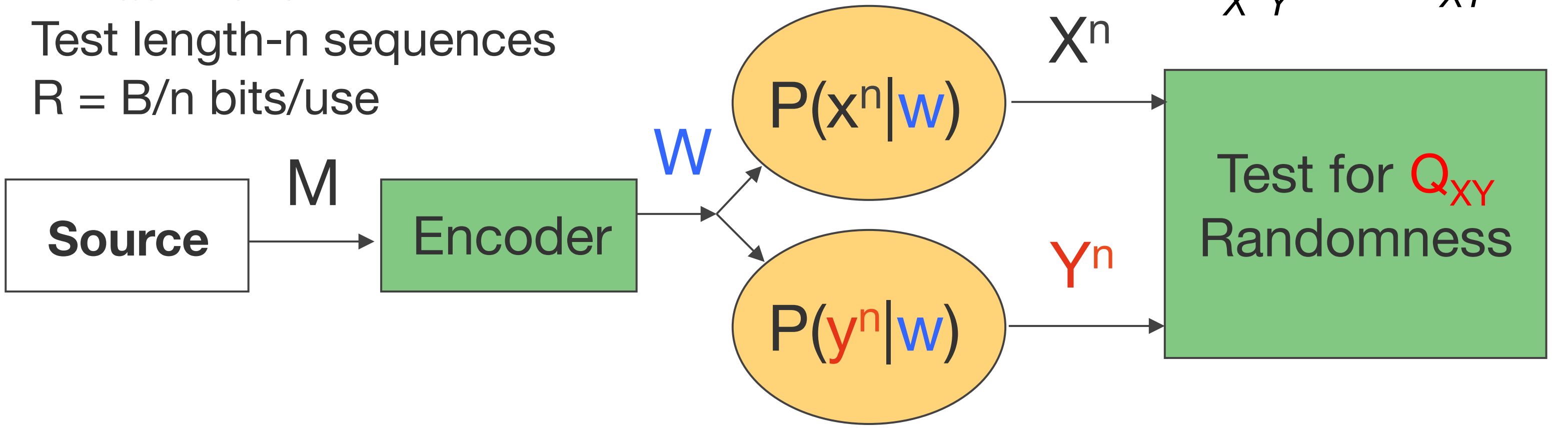
- Problem: find the maximum R for reliable communications: small Pr[M≠Ḿ]
- Random coding: choose each letter $x_i(m)$ independently via P_X
- Shannon's Capacity Function:

$$C = \max_{P_X} I(X;Y)$$

Common Information*

M has B bits
Test length-n sequences
R = B/n bits/use

$$P_{X^nY^n} \approx Q_{XY}^n = \prod_{i=1}^{n} Q_{XY}$$



- Problem: find the minimum R and channels $P(x^n|w)$, $P(y^n|w)$ so that

$$\frac{1}{n} D\left(P_{X^nY^n} \,\|\, Q_{XY}^n\right) \le \varepsilon \quad \text{for } \varepsilon > 0$$

- Result:

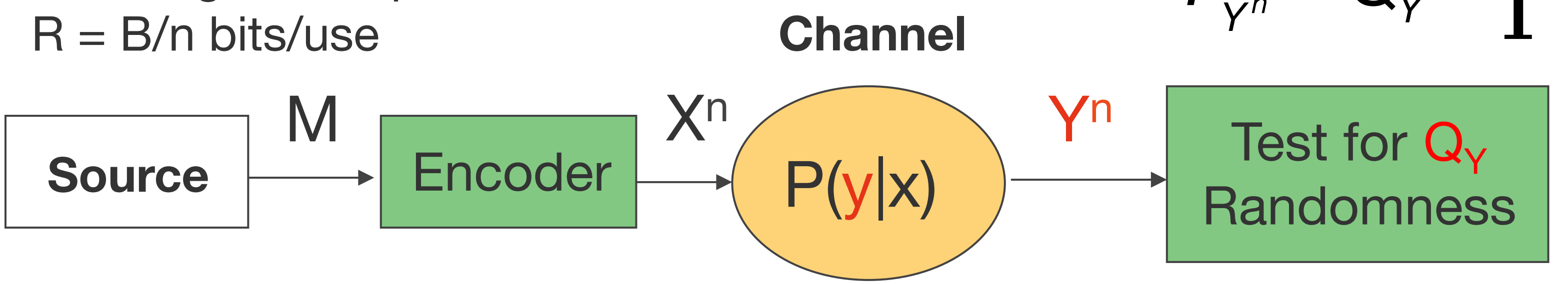
$$R = \min_{P_V P_{X|V} P_{Y|V}\,:\, P_{XY} = Q_{XY}} I(V;XY)$$

* Wyner 1975; above is the 2nd of Wyner's two approaches

Resolvability*

M has B bits
Test length-n sequences
R = B/n bits/use

$$P_{Y^n} \approx Q_Y^n = \prod_{i=1}^{n} Q_Y$$

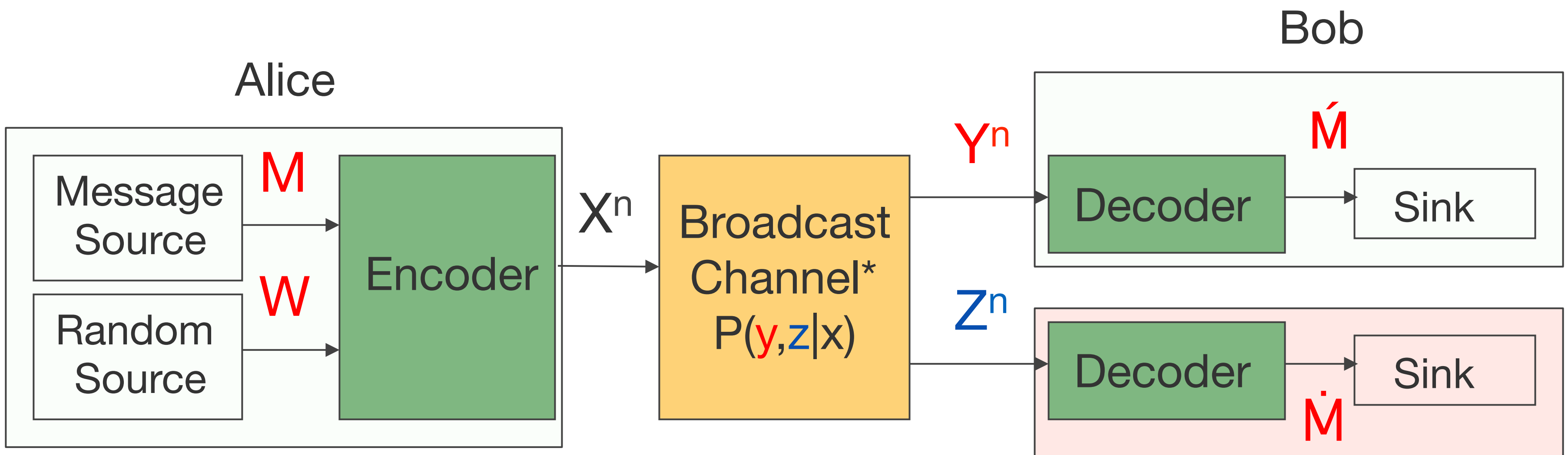


- Problem: find the minimum R so that $D\left(P_{Y^n} \| Q_Y^n\right) \leq \varepsilon$ for any $\varepsilon > 0$
- Random coding: choose each letter $x_i(m)$ independently via P_X
- Result:

$$R = \min_{P_X : P_Y = Q_Y} I(X;Y)$$

* Han-Verdú 1993 used variational distance $d_v=\|P_{Y^n}-P_Y^n\|_1$; For un-normalized divergence see, e.g., Winter 2005, Hayashi 2006, Watanabe-Oohama 2012, Hou-Kramer 2013

Wire-Tap Channel



M has B bits, $R = B/n$

W has B_1 bits, $R_1=B_1/n$

- Requirements: high rate R and
  - Reliability: error probability $P_e=\Pr[\acute{M}\neq M]$ should be small
  - Confusion/Secrecy: M should be "almost independent" of Z^n
  - Stealth/Covert: Z^n should "look like" a default Q_{Z^n}, typically an i.i.d. sequence of letters

Wyner 1975 (physically degraded BC: chain X-Y-Z is Markov); Csiszár-Körner 1978

Security Measures for Secrecy and Stealth

IT Security Measures

- Equivocation* (used by Wyner): $\Delta = \frac{1}{B} H(M \mid Z^n) = \frac{1}{nR} H(M \mid Z^n)$

  Goal: make Δ, 0 ≤ Δ ≤ 1, as large as possible
  Note: for Δ=1-ε get growing leakage Bε

- Alternatively: make 1-Δ as small as possible. If H(M)=B then

$$1 - \Delta = \frac{1}{B}\left(H(M) - H(M \mid Z^n)\right) = \frac{1}{B} I(M ; Z^n)$$

- Weak secrecy: $I(M;Z^n)/B$ or $I(M;Z^n)/n$

* to use unclear language to deceive

Security Measures (Continued)

- Weak secrecy: $I(M;Z^n)/B$ or $I(M;Z^n)/n$
- Criticism: if we fix the ratio then more bits leak as B grows. So perhaps we want an absolute measure.
- Strong secrecy*: $I(M;Z^n)$
- Remark: the approaches are effectively the same if we fix B
- Alternative**: measure variational distance $d_v = \left\| P_{MZ^n} - P_M P_{Z^n} \right\|_1$ and use** (B≥2, say d_v decreases faster than 1/B)

$$\frac{d_v^2}{2\ln 2} \le I(M;Z^n) \le d_v \log_2 \frac{2^B}{d_v}$$

- Most IT (and CS) papers since 1993 use d_v rather than $I(M;Z^n)$, which is somewhat strange

* Maurer 1993; Ahlswede-Csiszár 1993; ** Csiszár 1996

Security Measures (Continued)

- Strong secrecy: $I(M;Z^n) = D(P_{MZ^n} \| P_M P_{Z^n})$
- Stealth: $P_{Z^n} \approx Q_{Z^n}$ for some "default" Q_{Z^n}
- Effective secrecy*: replace the last P with Q

$$D(P_{MZ^n} \| P_M Q_{Z^n}) = \left\{ H(M) - \mathrm{E}\left[\log\left(Q_{Z^n}(Z^n)\right)\right]\right\} - H(MZ^n)$$

$$= I(M;Z^n) + D(P_{Z^n} \| Q_{Z^n})$$

- Remarks: (1) stronger than strong secrecy that has $Q_{Z^n} = P_{Z^n}$ (2) can study I & D separately; (3) “better” than var. distance; (4) we mainly study $Q_{Z^n} = Q_Z^n$; (5) worst case measures exist

* Independently used by Han-Endo-Sasaki 2013

Worst-Case Measures

- A natural worst-case* rather than an average metric is:

$$\max_m D(P_{Z^n|M=m} \| P_{Z^n}) \text{ rather than } I(M;Z^n) = D(P_{Z^n|M} \| P_{Z^n} \mid P_M)$$

- So a natural worst-case metric for us is (Q replaces P):

$$\max_m D(P_{Z^n|M=m} \| Q_{Z^n}) \text{ rather than } D(P_{Z^n|M} \| Q_{Z^n} \mid P_M)$$

- Remark: for design we wish to know how fast d_v or D approach zero with n, and not only the limit
- But we know that exponential dependence on n is possible
⇒ Should consider reasonable block length and code design

* Use standard expurgation arguments; valid for non-uniform M

Complexity-Based Security Measures

- Semantic Security* (Goldwasser & Micali 1984): based on Turing machines (other definitions: indistinguishability, non-malleability, non-dividability, etc.)
- Uses worst-case "advantage": consider g at Eve, h_r random

$$Adv = \max_{f,m}\left\{\max_{g} \Pr\left[g\left(Z^n\right) = f(m)\right] - \max_{h} \Pr\left[h_r\left(B\right) = f(m)\right]\right\}$$

* Wikipedia: A cryptosystem is semantically secure if any probabilistic, polynomial-time algorithm (PPTA) that is given the ciphertext of a certain message m (taken from any distribution of messages), and the message's length, cannot determine any partial information on the message with probability non-negligibly higher than all other PPTA's that only have access to the message length (and not the ciphertext)

Capacity

Capacity

- Result*:

$$C = \max_{P_{VX}:P_Z = Q_Z} \left[I(V;Y) - I(V;Z) \right]$$

  where chain V–X–YZ is Markov. The cardinality $|\mathcal{V}|$ is at most $|\mathcal{X}|$.
- Remarks:
  - C has same form as secrecy capacity except for the constraint
  - Stealth: if possible, choose Q_Z to maximize secrecy rate, i.e., as default send i.i.d X_i with P_X that maximizes the secrecy rate
  - Results extend to continuous-alphabet channels
  - Common complaint: C=0 if Bob's channel is worse than Eve's.
    How can we be sure this does not happen in practice?
    Reply 1: this can be reasonable
    Reply 2: the methods will improve security in any case

* Hou-Kramer 2013

- Further Remarks:
  - C depends on P(y|x) and P(z|x) only, not on "all" of P(y,z|x)
  - Physically degraded channel: chain X-Y-Z is Markov and thus

$$I(V;Y) - I(V;Z) = H(V|Z) - H(V|Y)$$
$$= I(V;Y|Z) \quad \ldots \text{ why?}$$
$$\leq I(X;Y|Z) \quad \ldots \text{ why?}$$
$$= I(X;Y) - I(X;Z) \quad \ldots \text{ why?}$$

  - Implication: best V is X
  - Stochastically degraded channel has P(y,z|x) where P(y|x) and P(z|x) are those of a physically degraded channel
  - Implications: same capacity C, and the best V is X

Example: BSC

$$C = \max_{P_X : P_Z = Q_Z} \left[I(X;Y) - I(X;Z) \right]$$

- BSCs: $Y = X \oplus A_1, \quad Z = X \oplus A_2$

  where $\Pr[A_1=1]=p_1$, $\Pr[A_2=1]=p_2$, $p_1 \le p_2 < 0.5$
- Channel is stochastically degraded (why?) so that best V is X
- Stealth: suppose we require $Q_Z(1)=q$ where $p_2 \le q \le (1-p_2)$
  We have* (try q=1/2 and $q=p_2$):

$$q = P_Z(1) = \left(1 - P_X(1)\right) p_2 + P_X(1)(1 - p_2) \;\Rightarrow\; P_X(1) = \frac{q - p_2}{1 - 2p_2}$$

$$C = H_2(p_2) - H_2(p_1) - H_2(q) + H_2\left((q - p_2)\frac{1 - 2p_1}{1 - 2p_2} + p_1 \right)$$

* $H_2(p) = -p \log_2 p - (1-p)\log_2(1-p)$

Example: AWGN Channel

$$C = \max_{P_X : P_Z = Q_Z} \left[I(X;Y) - I(X;Z) \right]$$

- AWGN Channels: $Y = X + A_1, \quad Z = X + A_2$

  where $A_1 \sim \mathcal{N}(0,N_1)$, $A_2 \sim \mathcal{N}(0,N_2)$, $0 \le N_1 \le N_2$
- Channel is stochastically degraded (why?) so that best V is X
- Stealth: suppose we require $Z \sim \mathcal{N}(0,Q)$ where $N_2 \le Q \le P+N_2$
  We have $X \sim \mathcal{N}(0,Q-N_2)$ and

$$C = \frac{1}{2}\log\left(1 + \frac{Q - N_2}{N_1}\right) - \frac{1}{2}\log\left(\frac{Q}{N_2}\right)$$

- Secrecy and covert capacities: $Q=P+N_2$ and $Q=N_2$, respectively

Proofs

Warning: lots of equations!

Random Coding Experiment

- Choose a P_X. Consider Shannon random coding experiment.
- Classic methods give $E[P_e|M=m,W=w] \to 0$ if $n \to \infty$ and

$$R + R_1 < I(X;Y)$$

- For secrecy & stealth, consider the following direct proof*:

$$D(P_{MZ^n|\mathrm{Code}} \| P_M Q_{Z^n}) = I(M;Z^n|\mathrm{Code}) + D(P_{Z^n|\mathrm{Code}} \| Q_{Z^n})$$

$$D\left(P_{Z^n|M=m,Code} \| Q_{Z^n}\right) = \sum_{w=1}^{2^{B_1}} \frac{1}{2^{B_1}} E\left[\log \frac{\sum_{j=1}^{2^{B_1}} P_{Z|X}^n\left(Z^n \middle| X^n(m,j)\right)}{2^{B_1} Q_{Z^n}\left(Z^n\right)} \middle| M=m, W=w\right]$$

* Hou-Kramer 2013; cf. Cuff 2009 and Yassaee 2013 who use concavity of x^2 for var. distance

- For a fixed z^n we have:

$$E\left[P^n_{Z|X}\left(z^n \middle| X^n(m,j)\right)\right] = P^n_Z\left(z^n\right)$$

- Using the concavity of log(.) and Jensen's inequality for the expectation over the code words $X^n(m,j)$ with $j \neq w$, we have

$$D\left(P_{Z^n|M=m,Code} \middle\| Q_{Z^n}\right)$$

$$\leq \sum_{w=1}^{2^{B_1}} \frac{1}{2^{B_1}} E\left[\log\left(\frac{P^n_{Z|X}\left(Z^n \middle| X^n(m,w)\right)}{2^{B_1} Q_{Z^n}\left(Z^n\right)} + \frac{P^n_Z\left(Z^n\right)}{Q_{Z^n}\left(Z^n\right)}\right) \middle| M = m, W = w\right]$$

- Alternatively, we have

$$D\left(P_{Z^n|M=m,Code} \| Q_{Z^n}\right) \le E\left[\log\left(\frac{P^n_{Z|X}\left(Z^n \middle| X^n\right)}{2^{B_1} P^n_Z\left(Z^n\right)} + 1\right)\right] + D\left(P^n_Z \| Q_{Z^n}\right)$$

- Keeping only δ-typical sequences, we "basically" have

$$D\left(P_{Z^n|M=m,Code} \| Q_{Z^n}\right) \le \log\left(\frac{2^{-n(1-\delta)H(Z|X)}}{2^{B_1} 2^{-n(1+\delta)H(Z)}} + 1\right) + D\left(P^n_Z \| Q_{Z^n}\right)$$

- As long as $R_1 > I(X;Z)$ and $Q_{Z^n} = P^n_Z$, avg. divergence is small

Coding Theorem

- Resulting rate bounds:

$$R + R_1 < I(X;Y) \quad \text{for reliability}$$

$$R_1 > I(X;Z) \qquad \text{for resolvability}$$

which gives:

$$R < I(X;Y) - I(X;Z)$$

- To get capacity:
  - replace X with V and generate code words $V^n(m,w)$
  - For each $V^n(m,w)$ generate $X^n(m,w)$ via artificial channel* $P_{X|V}$
- Default behavior for stealth: send i.i.d X_i with distribution P_X

* To reduce # random bits to $n \cdot I(X;Z)$: see Chia-El Gamal 2012 & Watanabe-Oohama 2015

Stealth Converse

- Several steps: $\xi \geq D\left(P_{MZ^n} \middle\| P_M Q_Z^n\right) = D\left(P_{Z^n|M} \middle\| Q_Z^n \middle| P_M\right)$

$$= \left[\sum_{z^n} P\left(z^n\right) \sum_{i=1}^{n} \log \frac{1}{Q_Z(z)}\right] - H\left(Z^n \middle| M\right)$$

$$\geq \sum_{i=1}^{n} \left[\sum_{z^n} P_{Z_i}(z) \log \frac{1}{Q_Z(z)}\right] - H(Z_i)$$

$$= \sum_{i=1}^{n} D\left(P_{Z_i} \middle\| Q_Z\right)$$

$$\geq nD\left(P_{Z_T} \middle\| Q_Z\right) \text{ where } P_T(i) = \frac{1}{n}, \quad i = 1, 2, \ldots, n$$

Secrecy Converse (Simplified)

- Main observation: can often replace 2 Csiszár sum identities steps with 1 telescoping identity
- As usual, Fano's inequality gives the first step

$$B = H(M) = I\left(M;Y^n\right) + H\left(M \middle| Y^n\right)$$

$$\leq I\left(M;Y^n\right) + \left(H_2\left(P_e\right) + P_e B\right)$$

- Requirement $I\left(M;Z^n\right) \leq \varepsilon n$ (weak secrecy) implies:

$$B \leq I\left(M;Y^n\right) + \left(\varepsilon n - I\left(M;Z^n\right)\right) + \left(H_2\left(P_e\right) + P_e B\right)$$

- Now use telescoping sum, set $U_i = Y^{i-1} Z_{i+1}^n$ and let T be a time-sharing RV and U_i-X_i-$Y_i Z_i$ forms a Markov chain for all i

$$
\begin{aligned}
& I(M;Y^n) - I(M;Z^n) \\
&= \sum_{i=1}^n \left[I(M;Y^i Z_{i+1}^n) - I(M;Y^{i-1} Z_i^n) \right] \quad (\textit{telescoping}) \\
&= \sum_{i=1}^n \left[I(M;Y_i \mid Y^{i-1} Z_{i+1}^n) - I(M;Z_i \mid Y^{i-1} Z_{i+1}^n) \right] \quad (\textit{chain rule}) \\
&= n\left[I(M;Y_T \mid U_T T) - I(M;Z_T \mid U_T T) \right]
\end{aligned}
$$

- Final steps*:

$$n\left[I\left(M;Y_T|U_T T\right)-I\left(M;Z_T|U_T T\right)\right]$$

$$\leq \max_{u} \max_{P_{MX|U}(\cdot|u)\in\Pi} n\left[I\left(M;Y|U=u\right)-I\left(M;Z|U=u\right)\right]$$

$$= \max_{P_{VX}\in\Pi} n\left[I\left(V;Y\right)-I\left(V;Z\right)\right]=nC$$

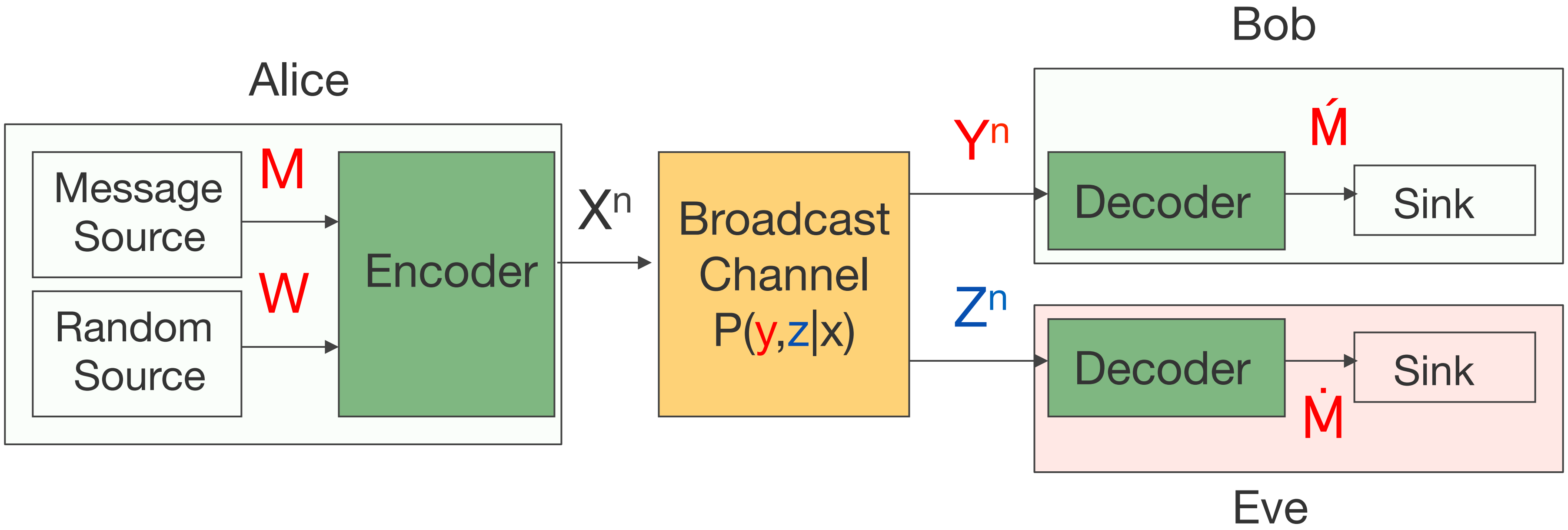
- Result with B=nR:

$$R \leq \frac{C+\varepsilon+H_2\left(P_e\right)/n}{1-P_e}$$

* Maximization constraint and cardinality bound follow by other steps

Operational Meaning of Stealth

Stealth and Binary Hypothesis Testing



- Since $D(P_{MZ^n} \| P_M Q_{Z^n}) = I(M;Z^n) + D(P_{Z^n} \| Q_{Z^n})$

  effective secrecy implies a small $D(P_{Z^n} \| Q_{Z^n})$
- Operational meaning? Can extend ideas from steganography*

* Cachin 2004

- Eve has two hypotheses:

$$H_0 : Q_{Z^n} \text{ (Alice transmits junk)}$$

$$H_1 : P_{Z^n} \text{ (Alice transmits information)}$$

- Error probabilities:

$$\alpha = \Pr\left[H_1 \text{ accepted } \middle| H_0 \text{ is true}\right] \text{ (false alarm)}$$

$$\beta = \Pr\left[H_0 \text{ accepted } \middle| H_1 \text{ is true}\right] \text{ (mis-detection)}$$

- Neyman-Pearson: test the ratio $Q_{Z^n}\left(z^n\right)/P_{Z^n}\left(z^n\right)$
- The set of z^n where H_0 is accepted:

$$A_F^n = \left\{ z^n : \frac{Q_{Z^n}\left(z^n\right)}{P_{Z^n}\left(z^n\right)} > F \right\}$$

- Error probabilities again:

$$\alpha = 1 - Q_{Z^n}\left(A_F^n\right) \quad \text{(false alarm)}$$

$$\beta = P_{Z^n}\left(A_F^n\right) \quad \text{(mis-detection)}$$

- Using Pinsker's inequality, we have

$$\sqrt{2\ln 2 \cdot D\left(P_{Z^n} \| Q_{Z^n}\right)} \geq \left\| P_{Z^n} - Q_{Z^n} \right\|_1$$

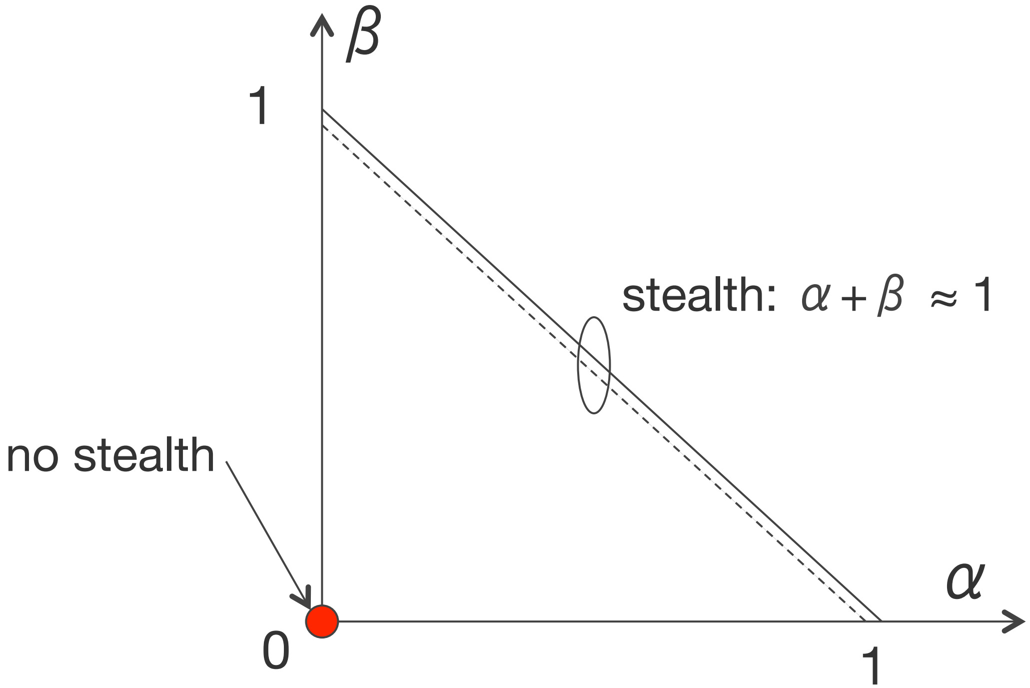
$$\geq \sum_{z^n \in A_F^n} \left| P_{Z^n}\left(z^n\right) - Q_{Z^n}\left(z^n\right) \right| \geq \left| \sum_{z^n \in A_F^n} P_{Z^n}\left(z^n\right) - Q_{Z^n}\left(z^n\right) \right|$$

$$\geq \left| P_{Z^n}\left(A_F^n\right) - Q_{Z^n}\left(A_F^n\right) \right| = \left| \beta - (1 - \alpha) \right|$$

- Thus, small $D(P_{Z^n} \| Q_{Z^n})$ means small $|\beta - (1-\alpha)|$ or

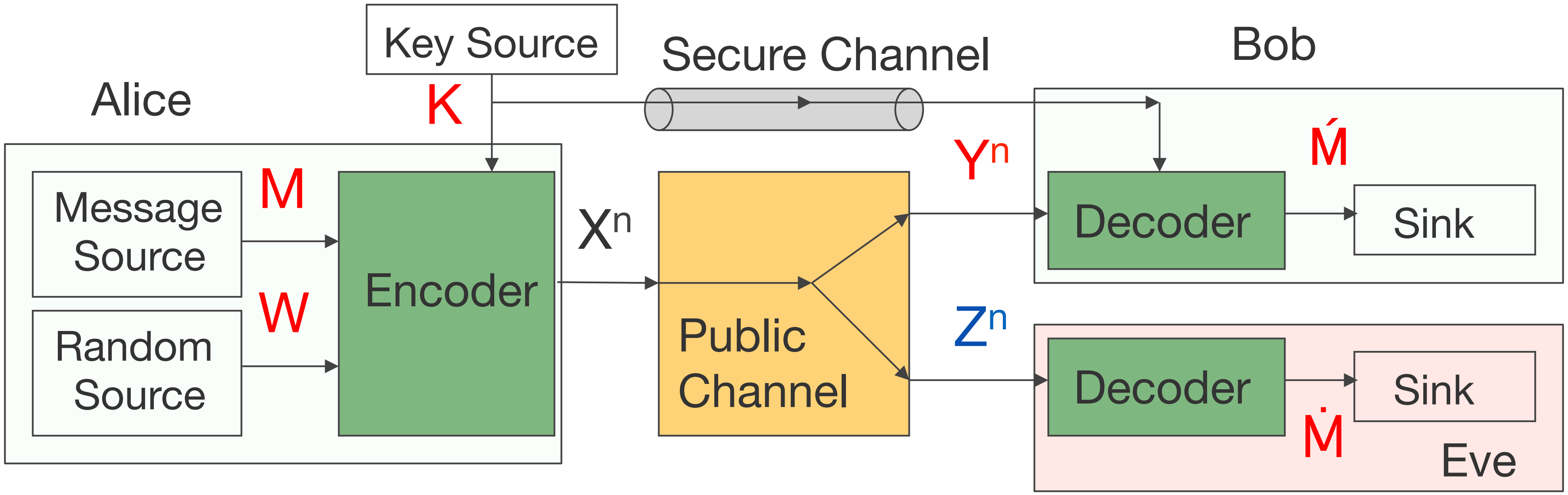
$$\alpha + \beta \approx 1$$

- But then Eve may as well guess without observing Z^n



Other Stealth/Covert Models

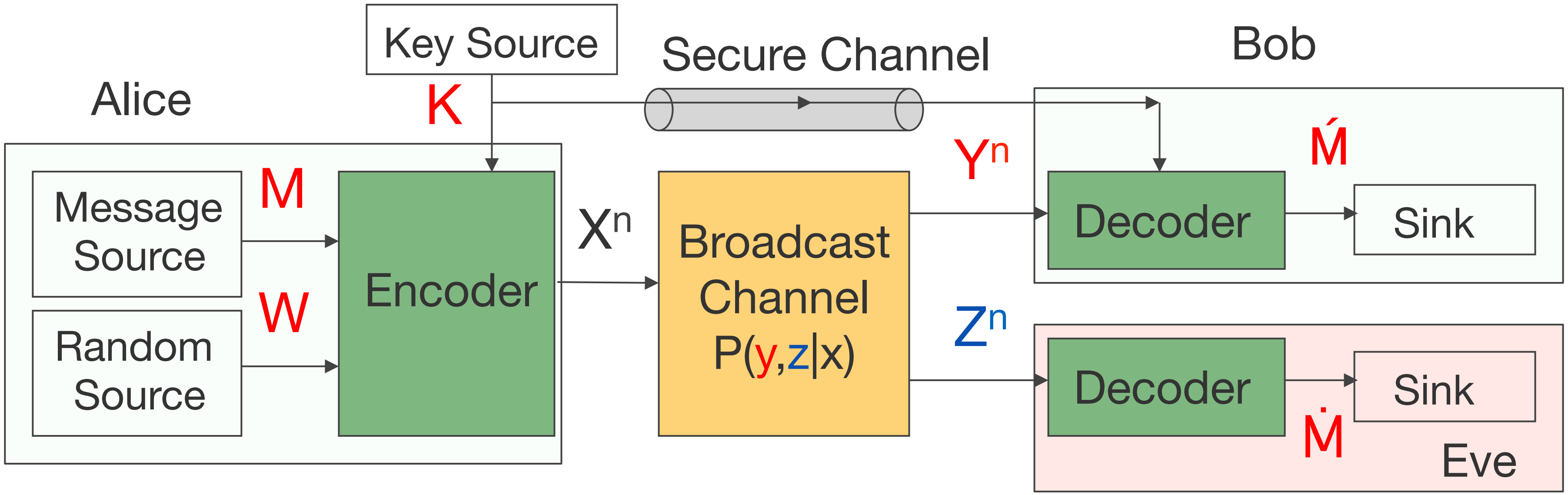
IT Stealth/Covert Models: An Incomplete History



- An IT steganography model*: Alice sends either (1) embedded message M=E via stegotext $X^n=S^n$ or (2) an open message via covertext $X^n=C^n$
- Require: (1) encoder does not know P_E (universality); (2) $I(\hat{E};E)>0$; (3) Bob knows when Alice is active; (4) secrecy via one-time pad**
- Limitations: (1) measure stealth via normalized divergence $D(P_{C^n} \| P_{S^n})/n$ (2) Universality only if H(E) is below threshold and rate loss is permitted

* Cachin 2004; ** secrecy for free

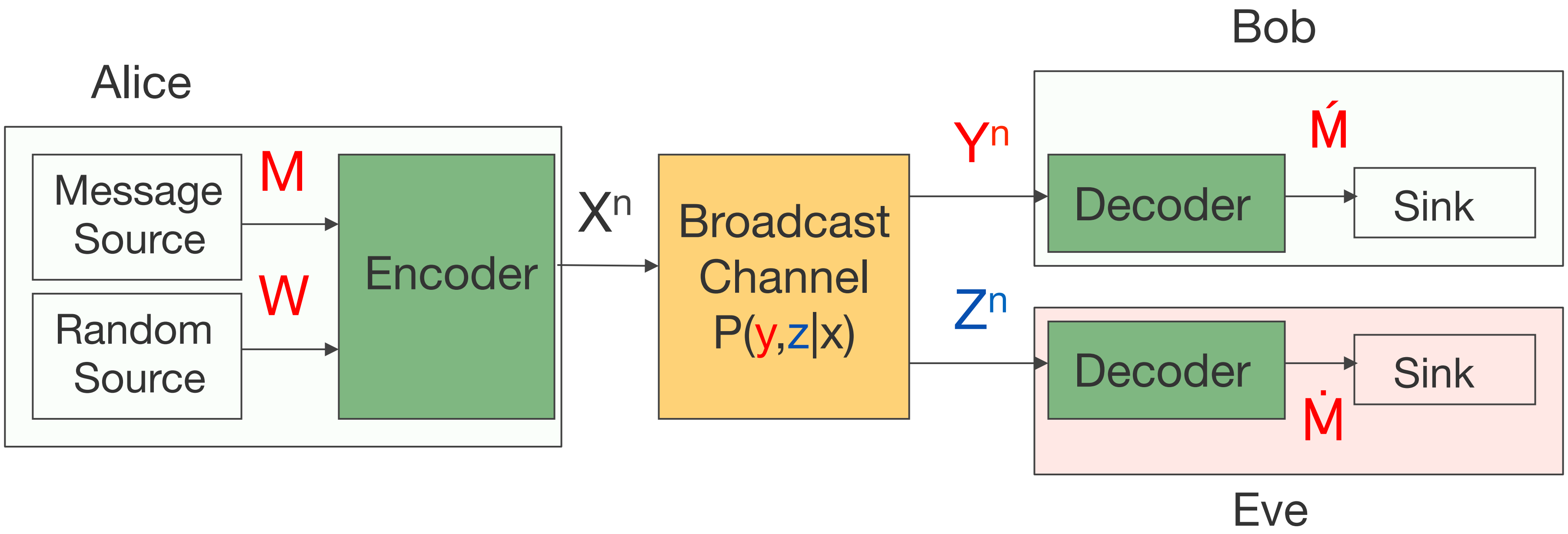
Recently



- Low probabability of detection (LPD)*: AWGN channel, Q_{Z^n} chosen for $X^n=0^n$
- Secret key: $B_K \sim n^{1/2}\log(n)$ bits* ... in fact, $n\cdot[I(X;Y)-I(X;Z)] \leq c\cdot n^{1/2}$ bits suffice
- Measure stealth via un-normalized $D(P_{Z^n} \| Q_{Z^n})$; note swap of cover/stegotext
- Result*: a square-root law due to local quadratic nature of divergence
- Result**: for BSCs, no need for K if Bob has a better channel than Eve (i.e., nI(X;Z)"deniability") but rate depends on channel differences

* Bash-Goeckel-Towsley 2012; ** Che-Bakshi-Jaggi 2013

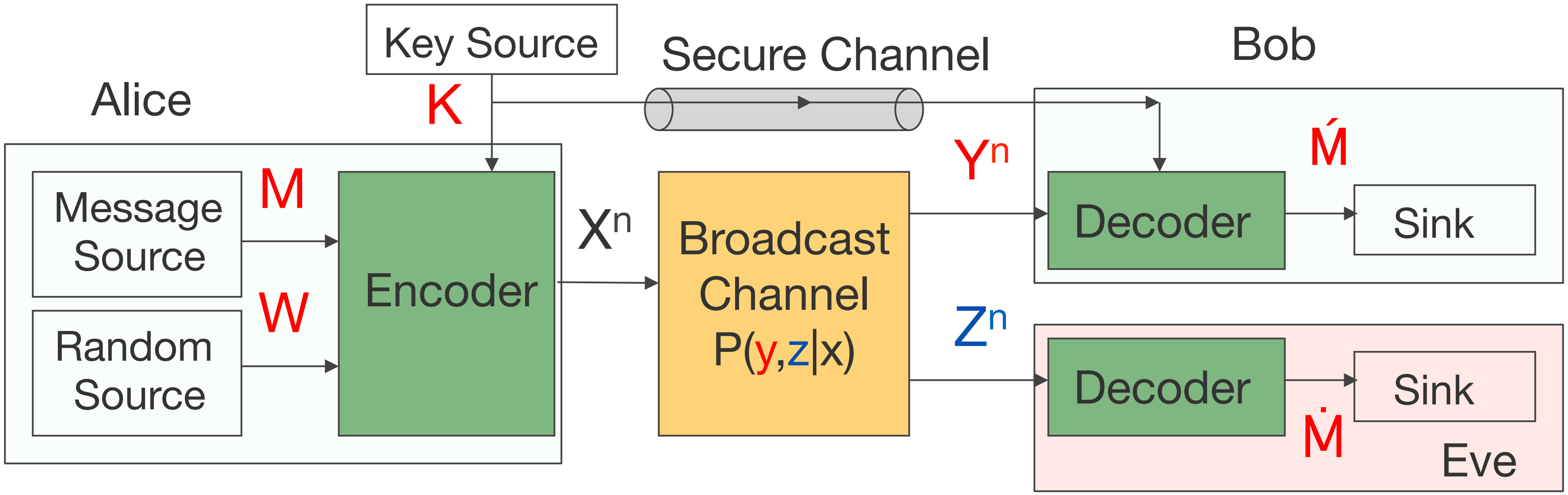
More Recently (up to end 2014)



- Consider (1) reliability; (2) secrecy; (3) stealth at the same time*
- Break* the square-root law if default (covertext) behavior is $X^n \neq 0^n$
- Other work: (1**) BSCs, variational distance, weak secrecy ($n^{1/2}$ normalization) (2***) noiseless compound channels; a “hidability” secrecy criterion uses probability ratios (worst case analysis similar to semantic security)

* Hou-Kramer 2013; Hou Dr. Ing. Thesis 2014;
** Che-Bakshi-Chan-Jaggi 2014; ***Kadhe-Bakshi-Jaggi-Sprintson 2014

Extensions



- Input* and output cost constraints (*Han-Endo-Sasaki 2013)
- Broadcast channel with a confidential message: add common message
- Secret key K with key rate** R_K ... security even if Ross has a better channel:

$$C = \max_{P_{VX} : P_Z = Q_Z} \left[I(V;Y) - \max\left(0, I(V;Z) - R_K\right) \right]$$

** replace W bits with K bits; leads to at most $n^{1/2}$ bits for LPD

Summary

Effective secrecy*

- includes the notion of stealth/covert communication;
- proofs use simple steps only

For more information, please see

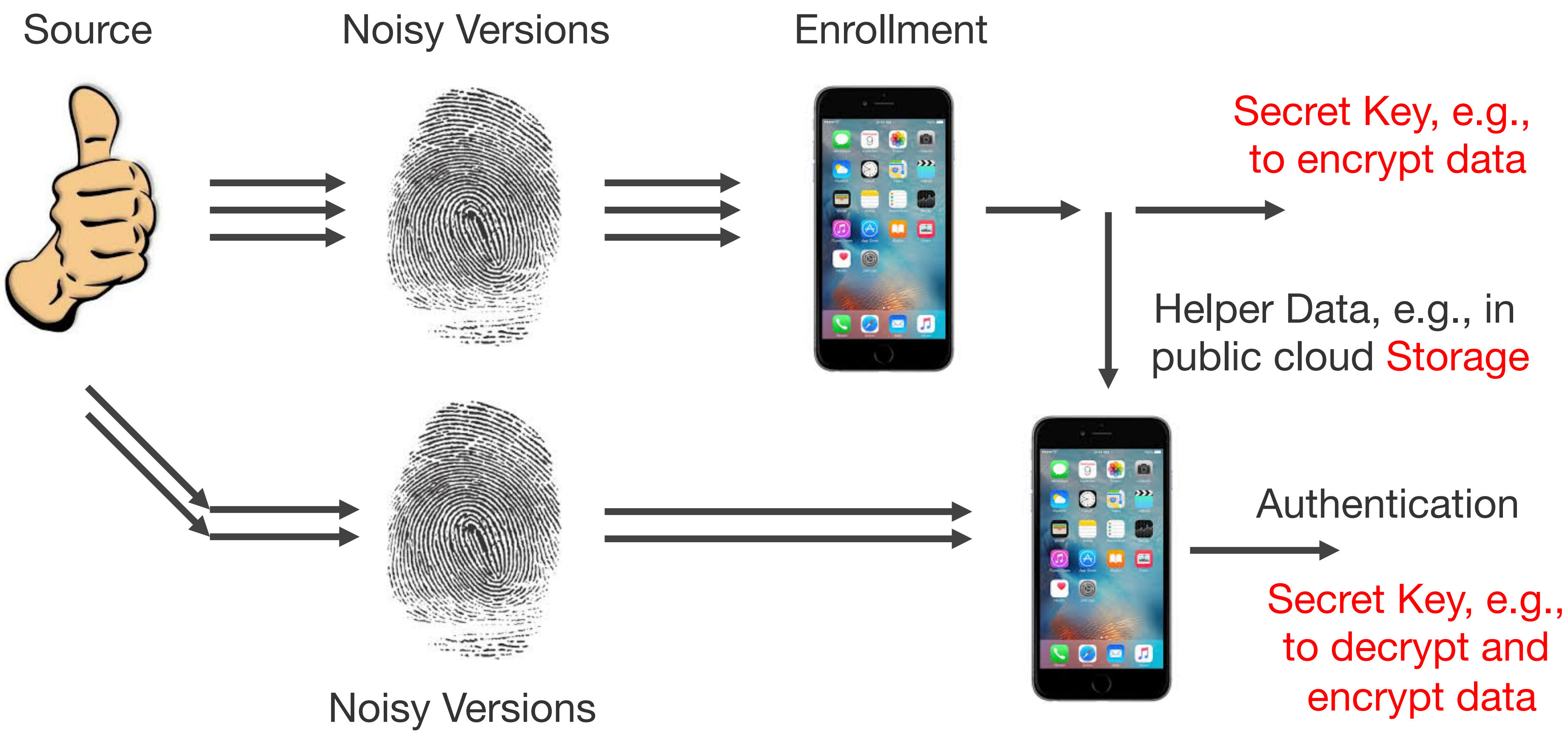
- J. Hou, “Coding for Relay Networks and Effective Secrecy for Wire-tap Channels”, Dr. Ing. Dissertation, TUM, Germany, 2014
- J. Hou and G. Kramer, “Effective secrecy: reliability, confusion and stealth,” arXiv:1311.1411, 2013 and 2014

* Independently introduced by Han-Endo-Sasaki 2013

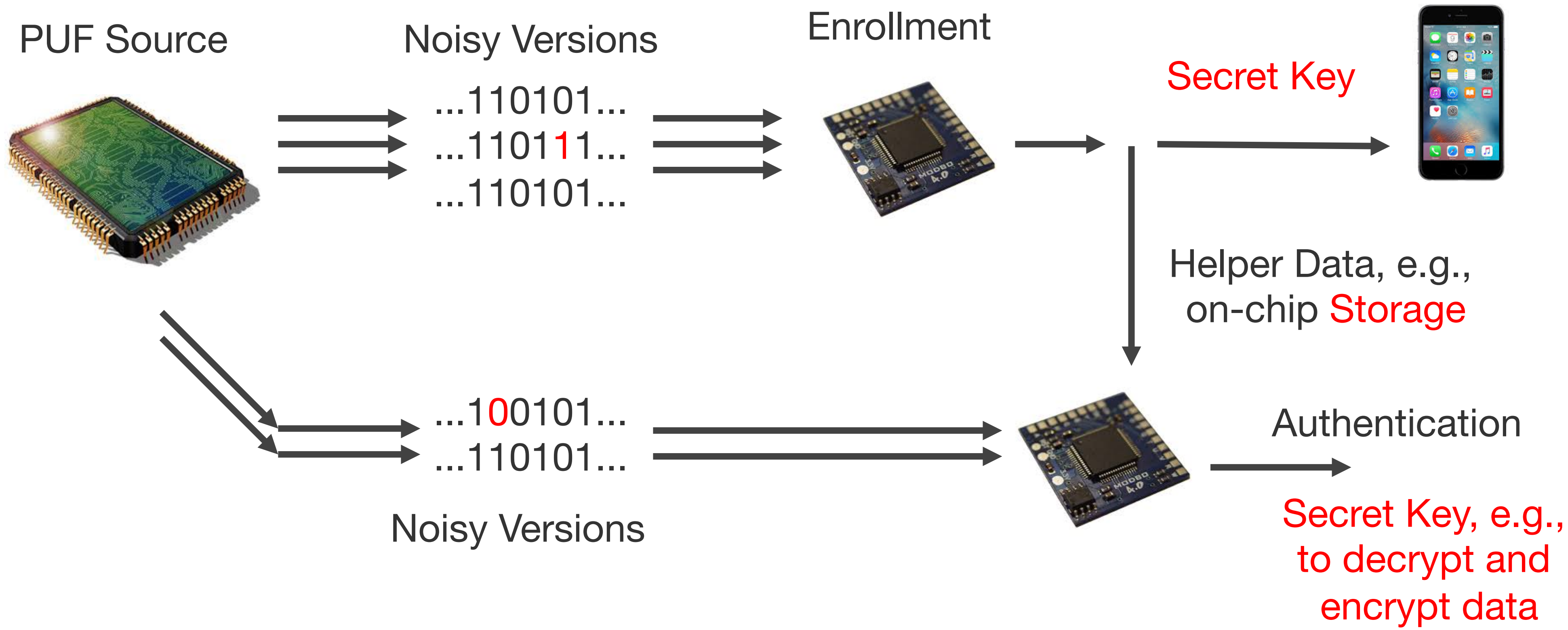
Part 2:
Secrecy, Privacy, and Storage for Noisy Identifiers

Motivation (Again) and Model

- Example A: Biometric Security

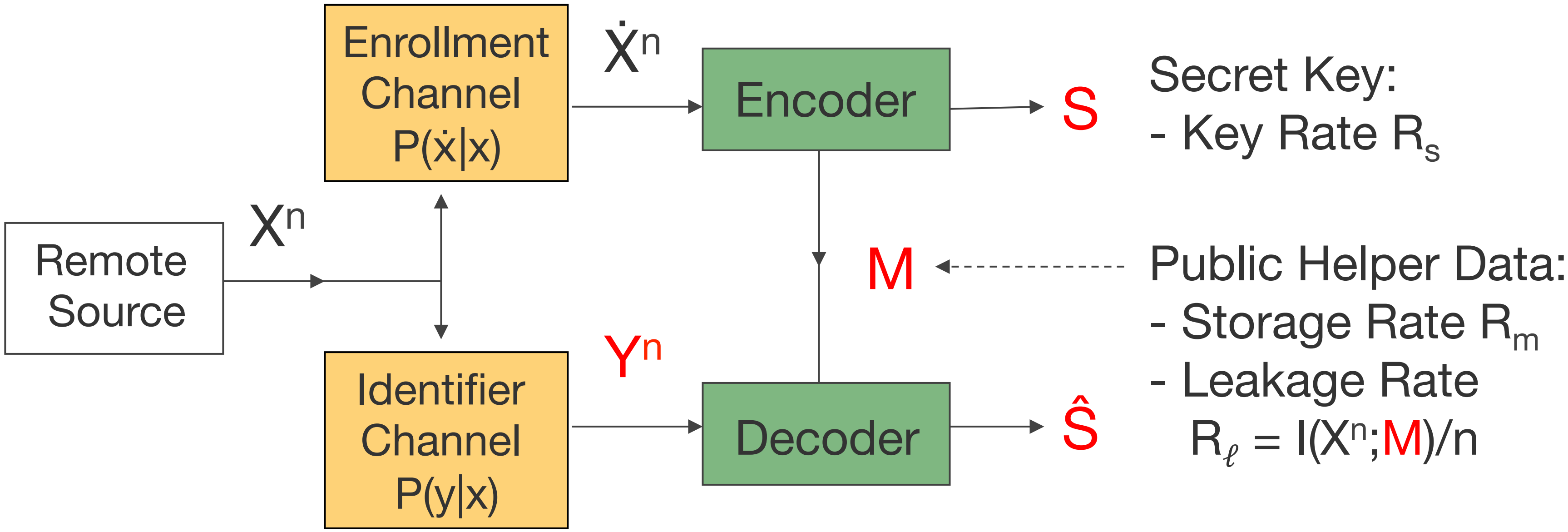


- Example B: Device Security for Things and their Internet, Hardware “Fingerprint” via a PUF*



* Physical Unclonable Function

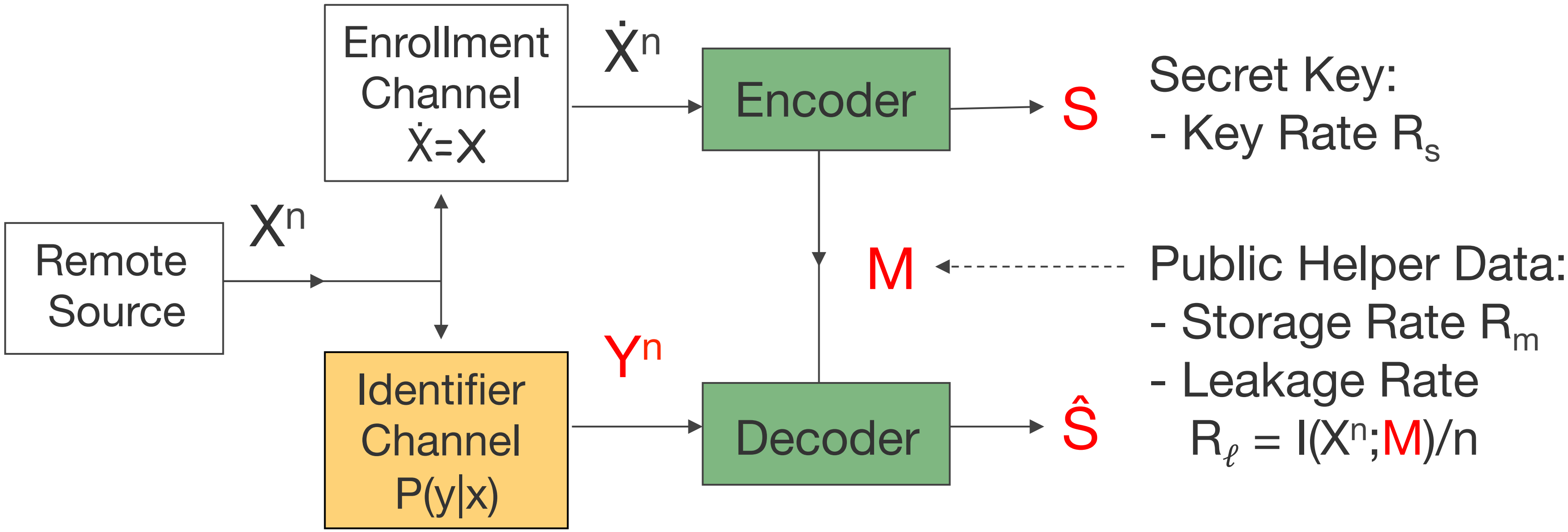
Noisy Identifier Model*



- Requirements:
  - Reliability: error probability $P_e = \Pr[\hat{S} \neq S]$ should be small
  - Secrecy: S should be independent of M and R_s large
  - Privacy: leakage rate R_ℓ should be small
  - Storage: storage rate R_m should be small

* Günlü-Kramer 2016

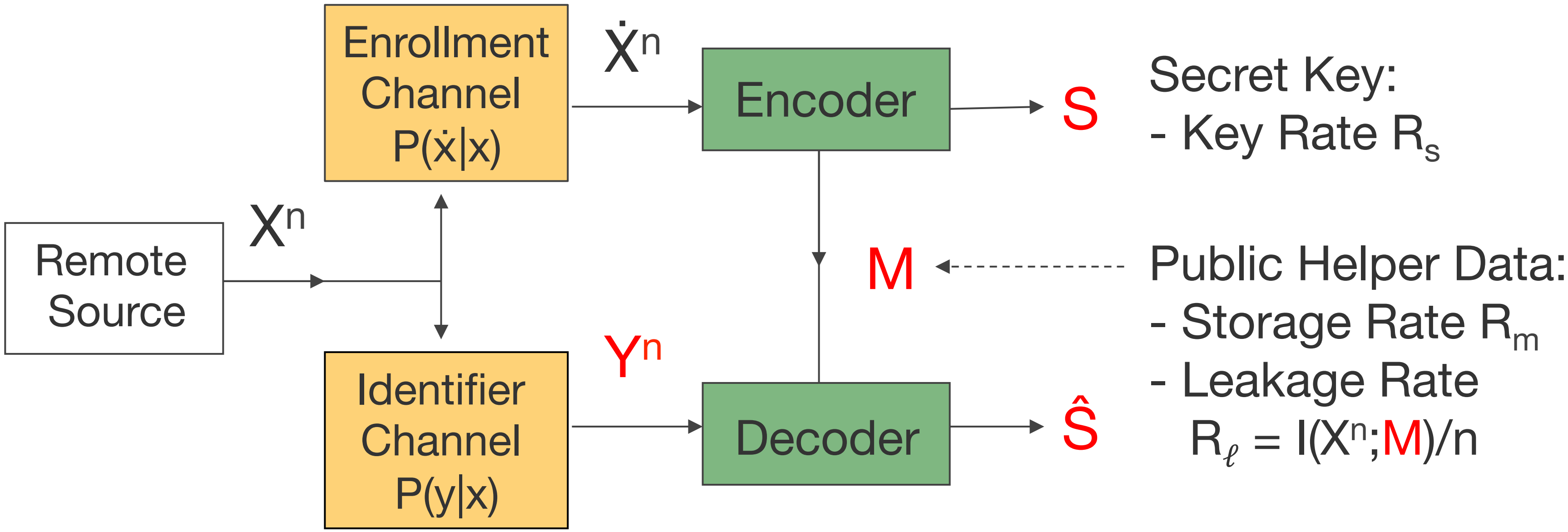
Variations (1)



- Remarks:
  - Commonly studied model* has noiseless enrollment: $\dot{X}=X$
  - Noisy identifier: R_s and R_m stay the same, R_ℓ decreases
  - So why study the noisy model? Two arguments:
    The correct model leads to practical insight and is a first step to model uncertainty about the source.

* Ignatenko-Willems 2009

Variations (2)



- Insights:
  - Multiple-measurements* have $\dot{X}$ and/or Y being vectors and multiple enrollment measurements are useful
  - Finite block-length results can be expected to lead to interesting tradeoffs

* Günlü-Kramer-Skórski 2015; Günlü-Kramer 2016

Security Measures and Capacity

Requirements

$$\Pr\left[S \neq \hat{S}\right] \leq \varepsilon$$

$$I(S;M)/n \leq \varepsilon$$

$$I\left(X^n;M\right)/n \leq R_\ell + \varepsilon$$

$$H(S)/n \geq R_s - \varepsilon$$

$$H(M)/n \leq R_m + \varepsilon$$

Remarks:

- ε small and positive
- Reliability
  Secrecy
  Privacy
  Key Rate
  Storage Rate

Capacity

$$\bigcup \left\{ \begin{array}{l} (R_s, R_\ell, R_m) : 0 \le R_s \le I(U;Y) \\ R_\ell \ge I(U;X) - I(U;Y) \\ R_m \ge I(U;\dot{X}) - I(U;Y) \end{array} \right\}$$

where union is over Markov chains U-Ẋ-X-Y

- Remarks:
  - If $\dot{X}=X$ then $R_\ell=R_m$ ... there are effectively two rates
  - Design for general Ẋ: one "simply" leaks less
  - Design for "wrong" X may violate requirements; a conservative approach designs for Ẋ (assuming model known)

Example: BSCs

- Noise-free enrollment: $\dot{X}=X$

$$\bigcup \left\{ \begin{array}{l} (R_s, R_\ell, R_m) : 0 \le R_s \le I(U;Y) \\ R_\ell = R_m \ge I(U;X) - I(U;Y) \end{array} \right\}$$

where union is over Markov chains U-X-Y

- BSC: Y = X+Z mod 2, Pr[X=1]=0.5, Pr[Z=1]=p, $0 \le p < 0.5$
- Problem: maximize I(U;Y) while minimizing I(U;X) aka the information bottleneck problem*
- Solved by using Mrs. Gerber's Lemma** which implies:

$$H(Y|U) \ge h\left(p * h^{-1}\left(H(X|U)\right)\right)$$

with equality if the U-to-X channel is a BSC with crossover probability $h^{-1}(H(X|U))$ ... here p*q denotes "cyclic convolution"

* Witsenhausen-Wyner 1975; Tishby-Pereira-Bialek 1999; ** Wyner-Ziv 1973

- Given U, let q=h^{-1}(H(X|U)). Mrs. Gerber's Lemma implies:

$$H(Y|U) \geq h(p * q)$$

- Proof steps:

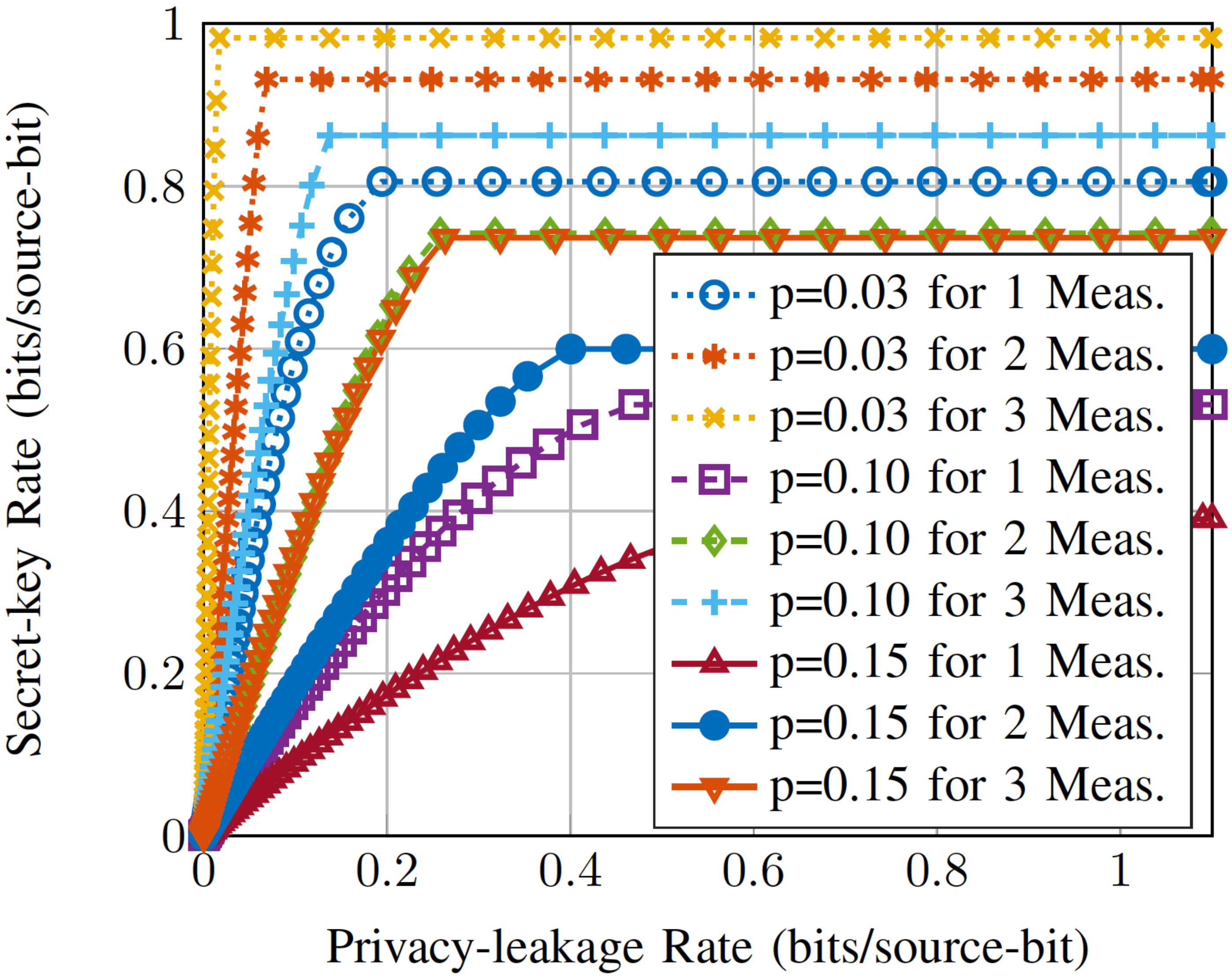
$$I(U;Y) = H(Y) - H(Y|U) \leq H(Y) - h(p * q)$$

$$\begin{aligned} I(U;X) - I(U;Y) &= H(X) - H(Y) + H(Y|U) - H(X|U) \\ &\geq H(X) - H(Y) + h(p * q) - H(X|U) \end{aligned}$$

- So a BSC from U-to-X is best: must optimize one number only
- Results extend (with a few limitations) to multiple measurements during enrollment and identification*

* Günlü-Kramer-Skórski 2015; Günlü-Kramer 2016

Example: BSCs and Multiple Measurements*



- R_s vs. R_ℓ=R_m
- Biometrics: low leakage R_ℓ
- PUFs: large key rate R_s and (then) minimal leakage rate R_ℓ

* Günlü-Kramer-Skórski 2015

Summary

Biometric and Device Security

- use unique variations to authenticate and produce keys
- measurement process is noisy: use error control codes
- three parameters: security, privacy, storage

For more information about PUFs, see

- "Physical Unclonable Functions and Applications: A Tutorial", Proc. IEEE, vol. 102, no. 8, 2014