Dependence in IV-related bytes of RC4 key enhances vulnerabilities in WPA

Sourav Sen Gupta[1] Subhamoy Maitra[1] Willi Meier[2]
Goutam Paul[1] Santanu Sarkar[3]

Indian Statistical Institute, India

FHNW, Windisch, Switzerland

Chennai Mathematical Institute, India

FSE 2014 London, 4 March 2014

RC4 and WPA

RC4 Stream Cipher

- Invented in 1987; simplest cipher to date.
- Several statistical weaknesses discovered.
- Still one of the most common ciphers in use.

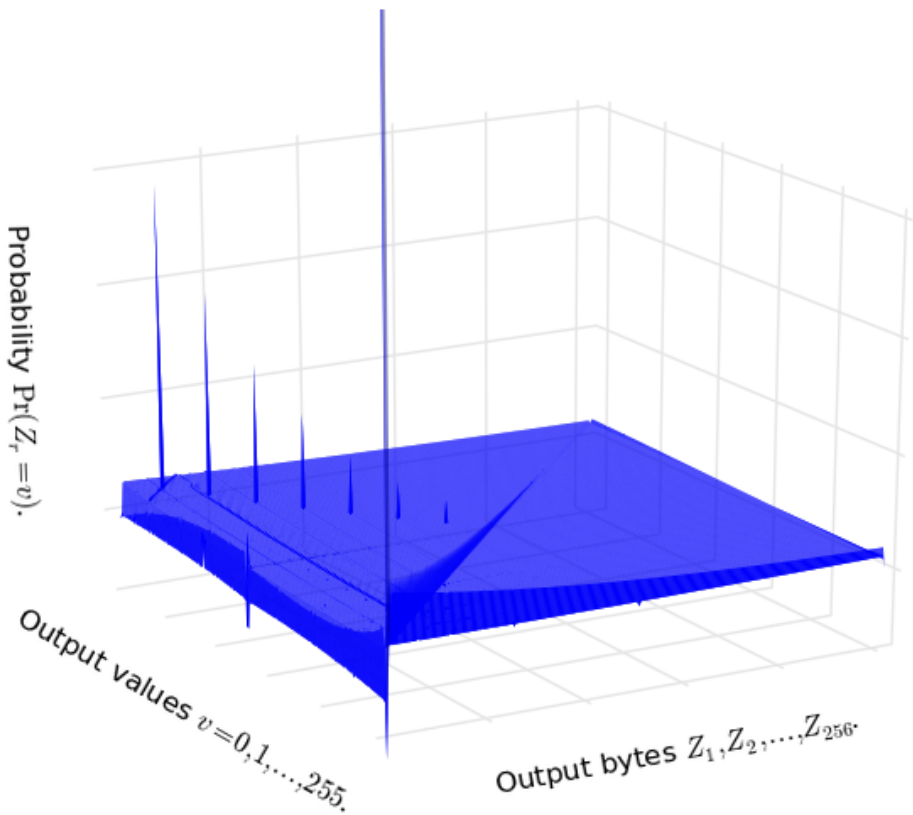
WPA Protocol

- Uses RC4 as the core cipher for encryption.
- Successor of WEP, which used RC4 as well.
- TKIP generates 16-byte RC4 key per frame.

Results on RC4

Statistical weaknesses in RC4

Significant biases in $Z_2 = 0$, $Z_1 = v$, $Z_r = 0$, $Z_r = r$, $Z_r = -r$.

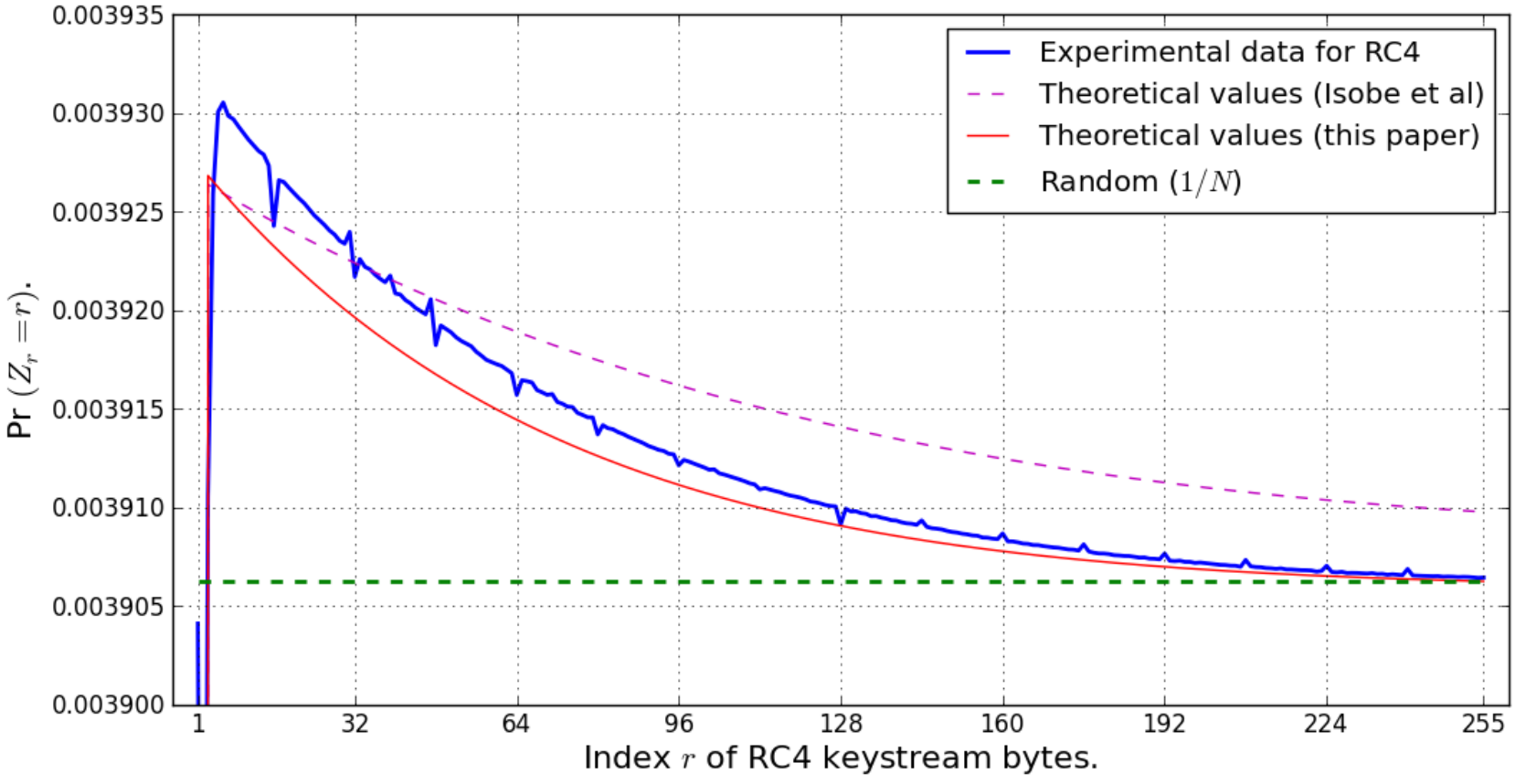


Data – AlFardan et al., USENIX 2013 – On the Security of RC4 in TLS and WPA (http://www.isg.rhul.ac.uk/tls/)

Statistical weaknesses in RC4

$Z_2 = 0$	observation and proof	Mantin and Shamir, 2001
$Z_1 = v$	observation	Mironov, 2002
	proof	Sen Gupta et al., 2012
$Z_r = 0$	observation and proof	Maitra et al., 2011
$Z_l = -l$	observation and proof	Sen Gupta et al., 2011-12
$Z_{xl} = -xl$	observation and proof	Isobe et al., 2013
$Z_r = r$	observation and proof	Isobe et al., 2013
	observation	AlFardan et al., 2013

Result 1 : Proof of $Z_r = r$



$$\Pr(Z_r = r) = \frac{1}{N} + \Pr(S_0[1] = r) \cdot \frac{1}{N}\left(1 - \frac{1}{N}\right)\left(1 - \frac{r-2}{N}\right)\left(1 - \frac{2}{N}\right)^{r-3}$$

Beyond the initial 255 bytes

- RC4 'recycles' after first 255 rounds
- We generally consider only up to initial 255 bytes
- General expectation – no significant bias after that

Recent results indicate otherwise

$Z_{256} = 0$	observation	Isobe et al., 2013
	observation	AlFardan et al., 2013
	proof	Sarkar et al., 2013
$Z_{257} = 0$	observation	Isobe et al., 2013
	proof	Sarkar et al., 2013

Result 2 : Bias in Z_{259}

Theorem

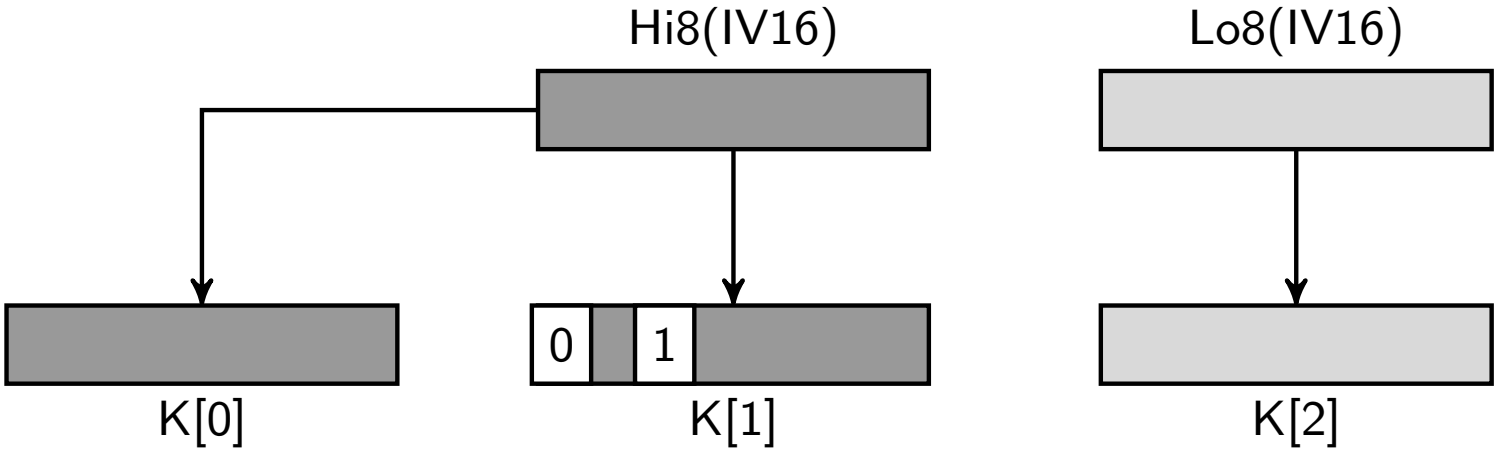
The probability that the $(N+3)$-th keystream byte of RC4 is 3 is

$$\Pr(Z_{N+3} = 3) \approx \frac{1}{N} + \frac{0.18}{N^2}.$$

Implication of this result – plaintext recovery attack on byte 259 may now use this single byte bias, instead of long-term biases.

Results on WPA

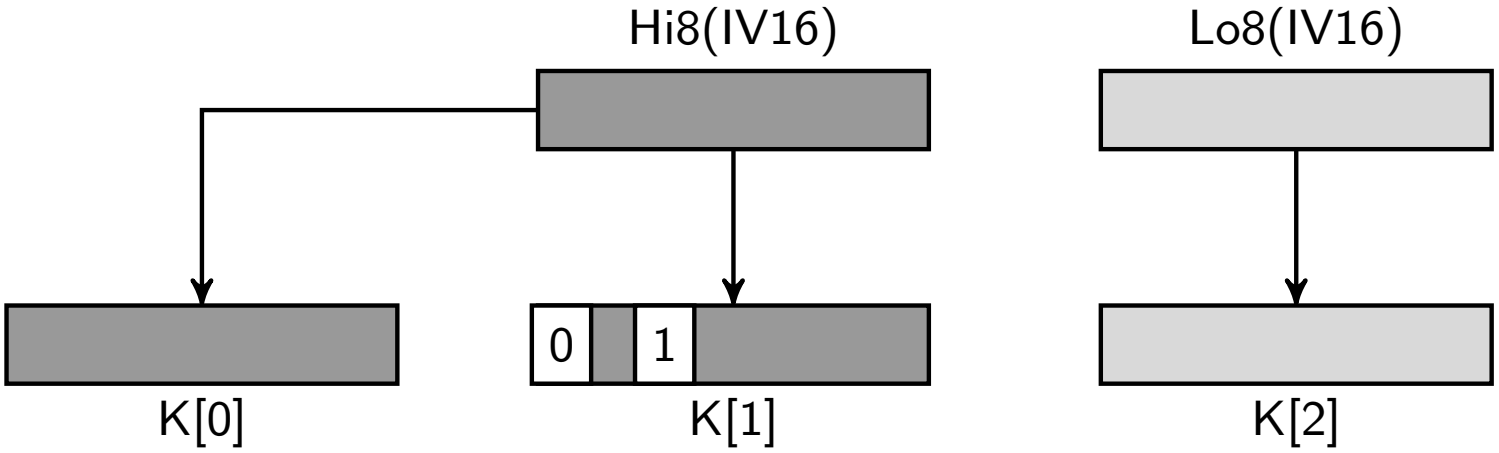
Motivation : IV-dependence in WPA



First three bytes of the 16-byte RC4 key of WPA/TKIP

$$K[0] = (\text{IV16} >> 8) \;\&\; \text{0xFF}$$
$$K[1] = ((\text{IV16} >> 8) \mid \text{0x20}) \;\&\; \text{0x7F}$$
$$K[2] = \text{IV16} \;\&\; \text{0xFF}$$

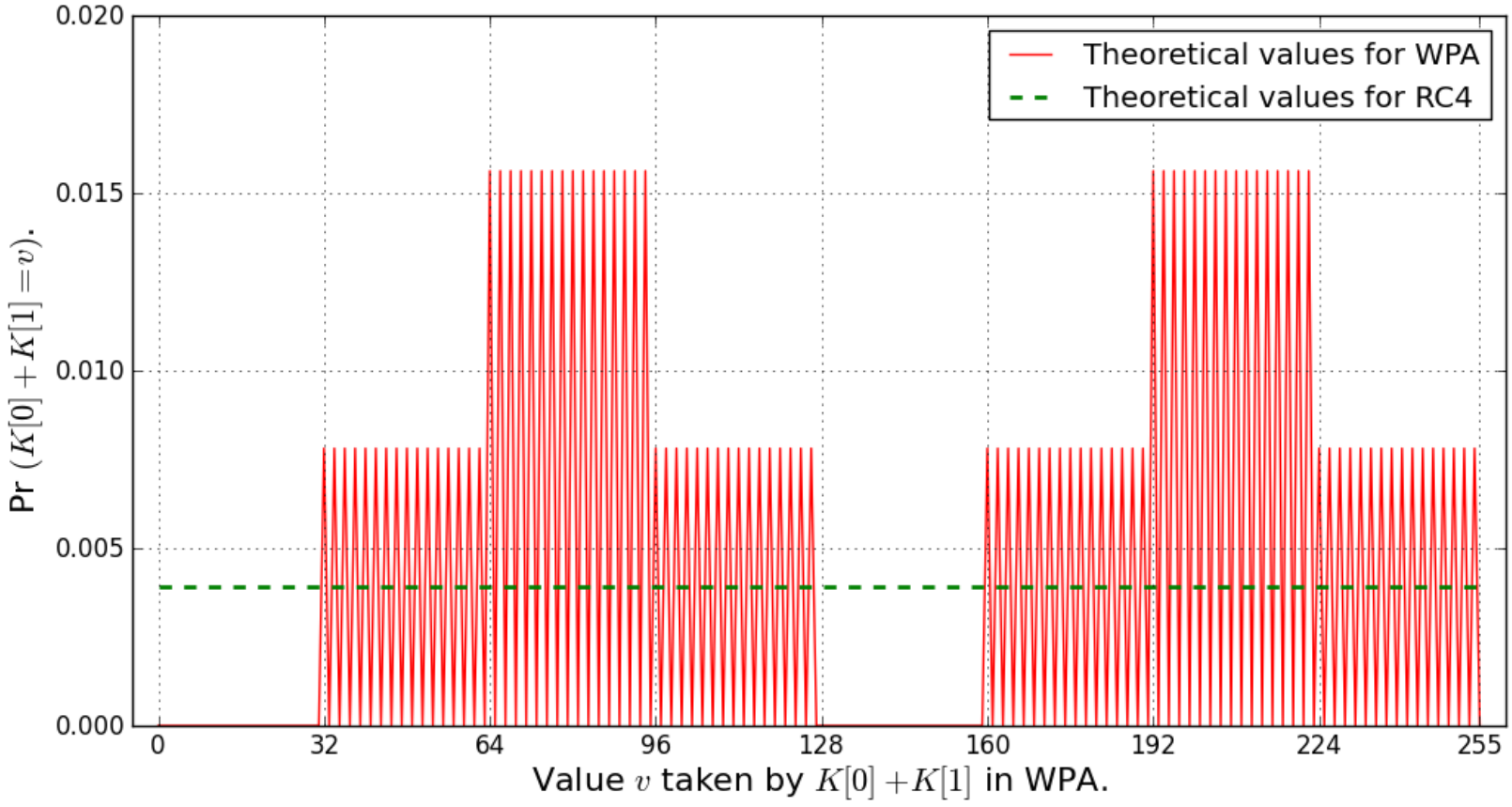
Motivation : IV-dependence in WPA



First two bytes of the 16-byte RC4 key of WPA/TKIP

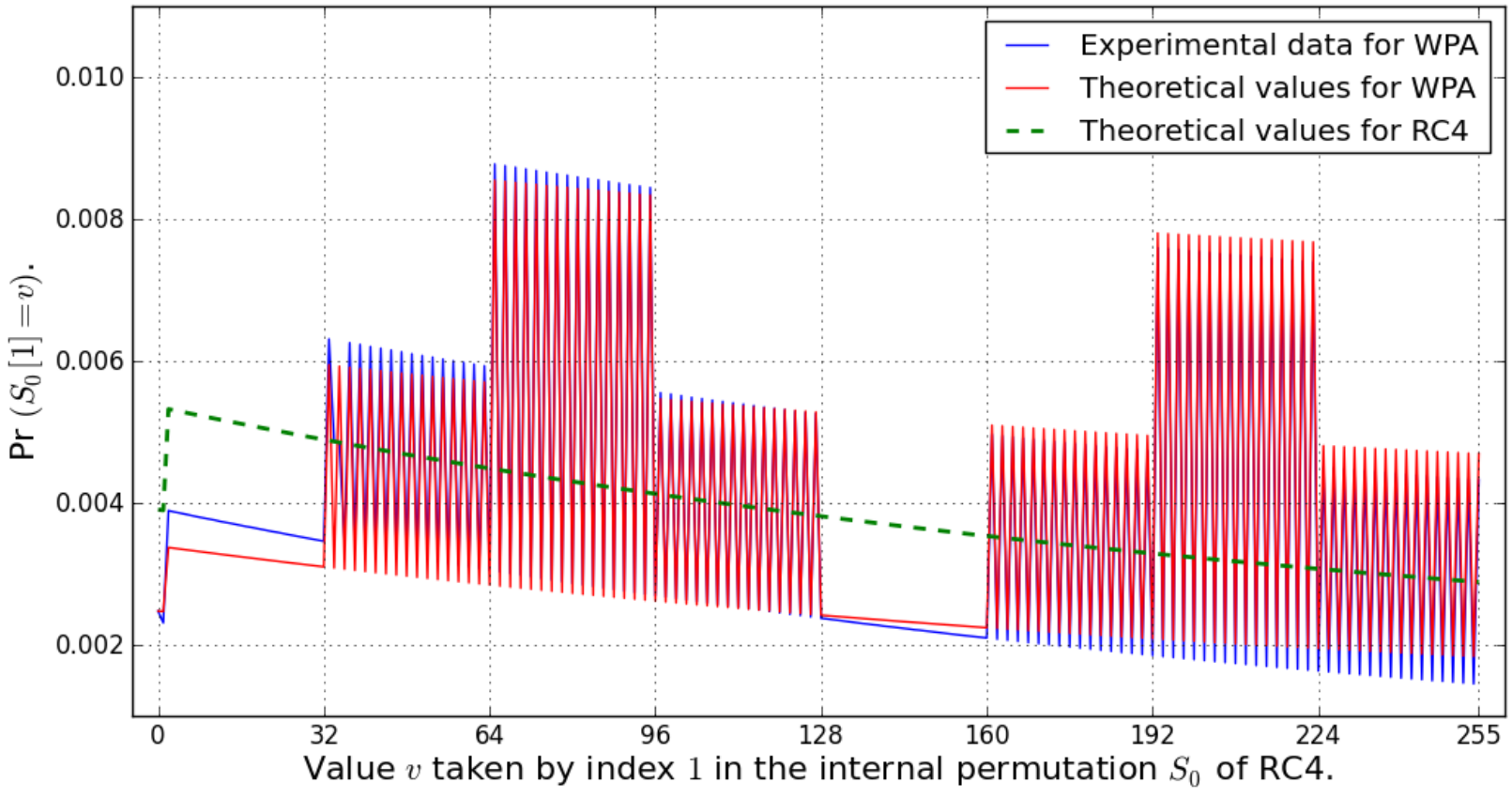
- $K[0]$ and $K[1]$ have at least 6 bits in common!
- $K[0] + K[1]$ is always even, and can't take all values either.

Observation : Distribution of $K[0] + K[1]$



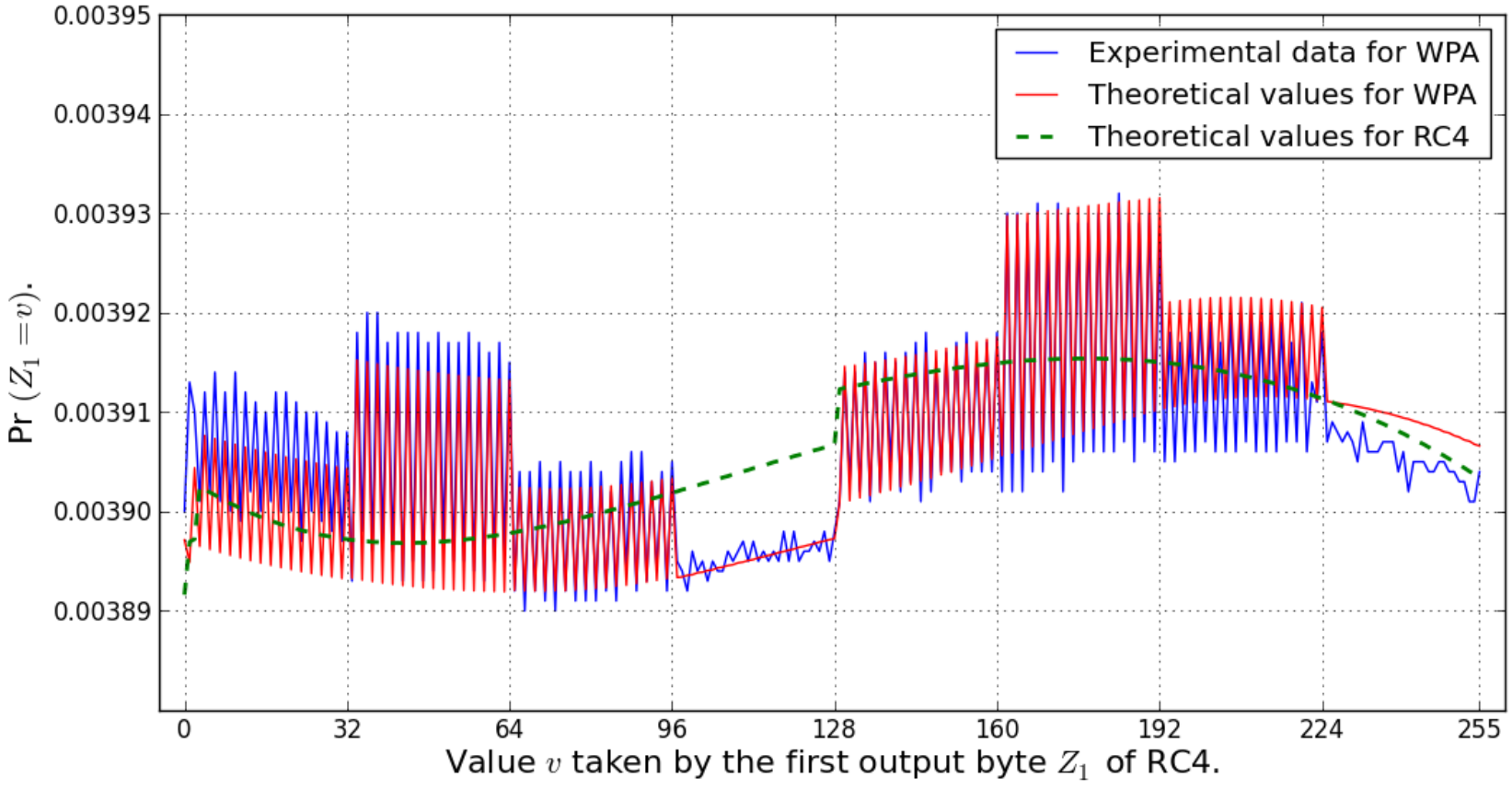
Known – Roos' bias : $S_0[1]$ is biased towards $K[0] + K[1] + 1$.

Result : $K[0] + K[1] \longrightarrow S_0[1]$



Known – Sen Gupta et al. : Distribution of Z_1 depends on $S_0[1]$.

Result : $K[0] + K[1] \longrightarrow S_0[1] \longrightarrow Z_1$



This proves the experimental observation by AlFardan et al., 2013.

WPA distinguisher based on Z_1

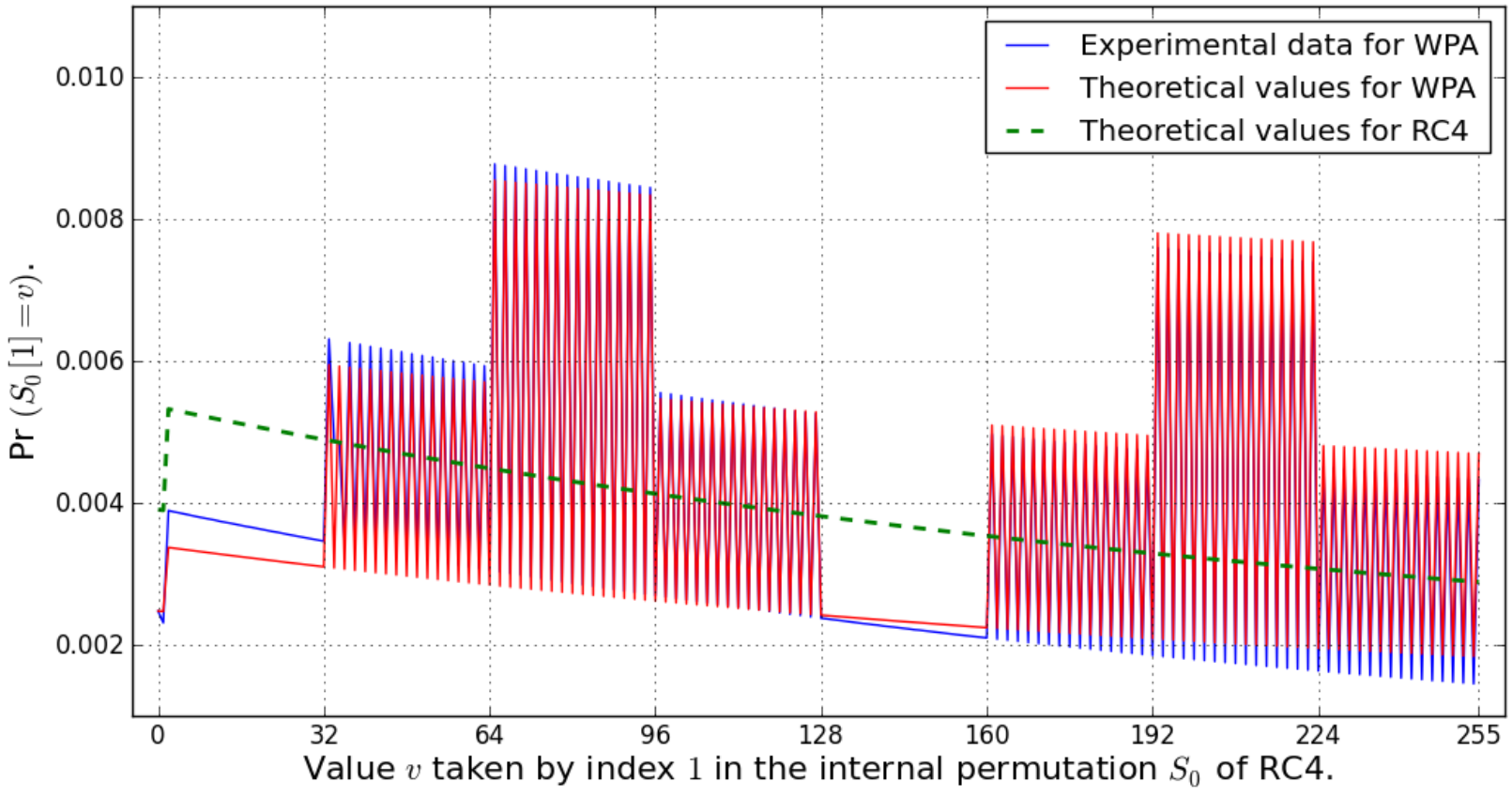
Event : Z_1 is even

- Probability in RC4 : $= 0.4999946 = p$
- Probability in WPA : $= 0.5007041 = p(1 + q)$
- Thus, $p = 0.4999946 \approx 1/2$ and $q \approx 0.001419 \approx 0.363/N$

Sample complexity : $1/pq^2 \approx 8N^2 = 2^{19}$ bytes.

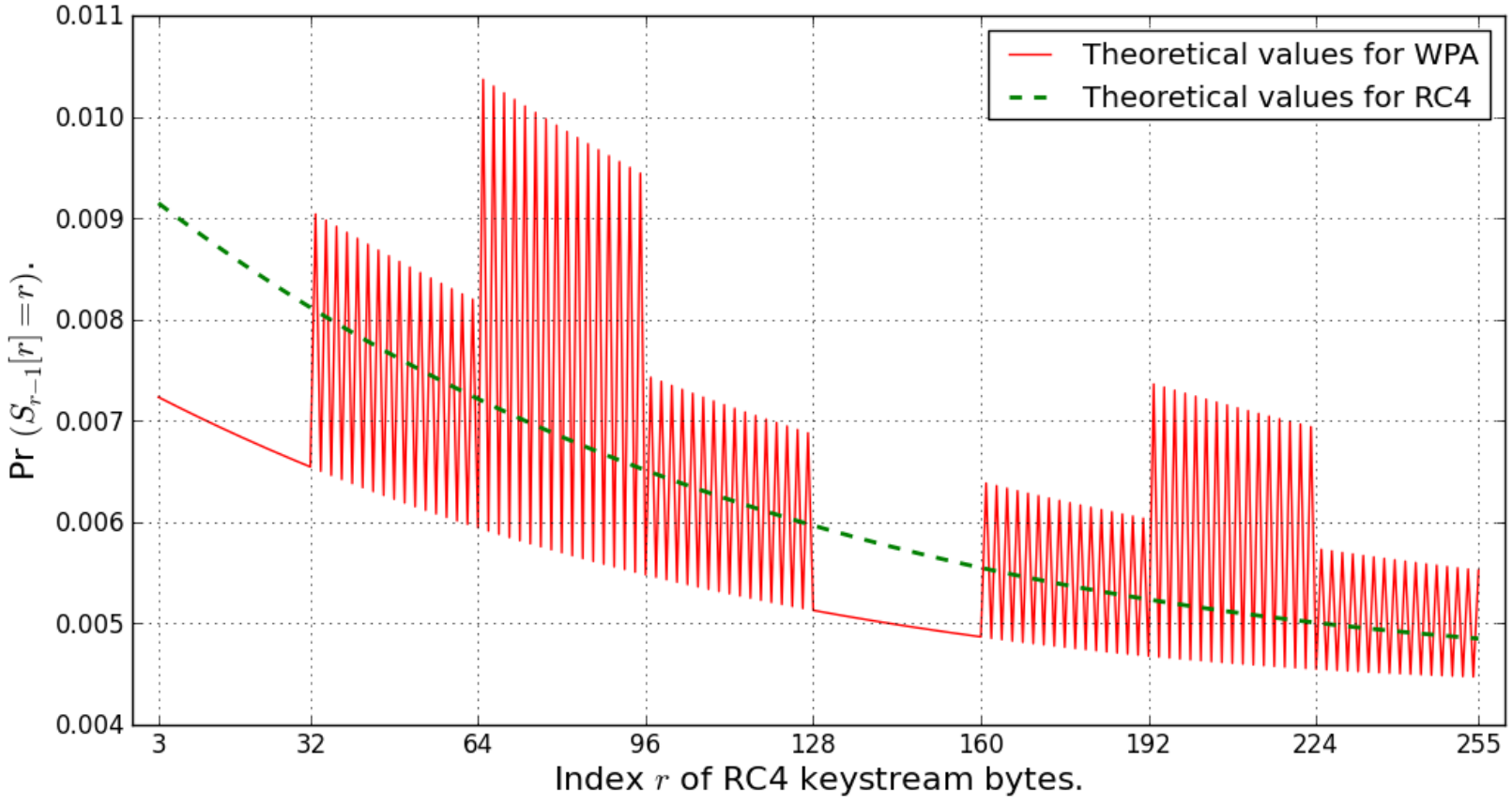
This result beats the best existing WPA distinguisher of Sepehrdad et al. (2011-12), which requires more than 2^{40} samples.

Recall : $K[0] + K[1] \longrightarrow S_0[1]$



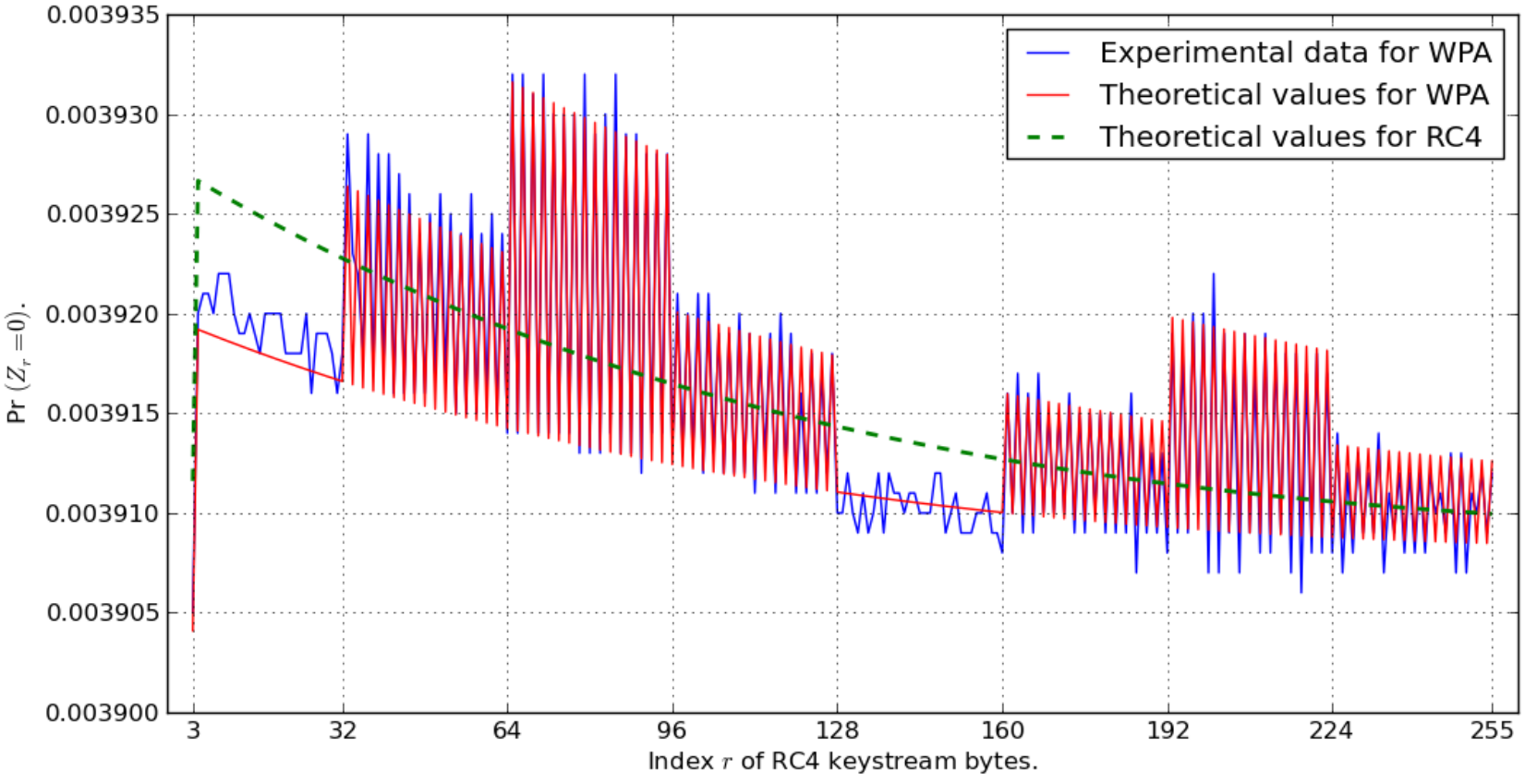
Known – Sen Gupta et al. : Distribution of $S_{r-1}[r]$ depends on S_0.

Result : $K[0] + K[1] \longrightarrow S_0[1] \longrightarrow S_{r-1}[r]$



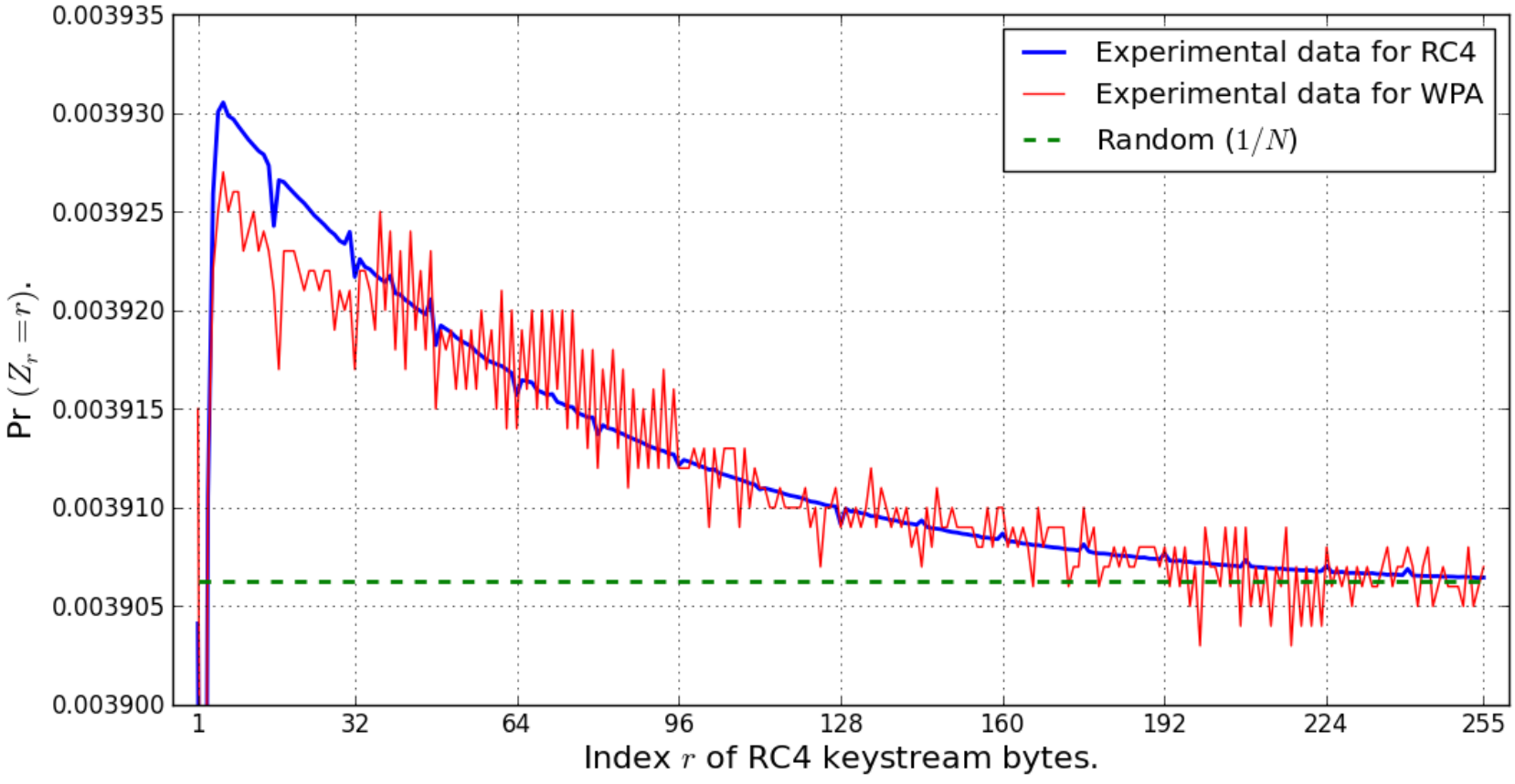
Known – Sen Gupta et al. : Distribution of Z_r depends on $S_{r-1}[r]$.

Result : $K[0] + K[1] \longrightarrow S_0[1] \longrightarrow S_{r-1}[r] \longrightarrow Z_r$



This proves the experimental observation by AlFardan et al., 2013.

Observation : Bias in $Z_r = r$



Intuition : $K[0] + K[1] \longrightarrow S_0[1] \longrightarrow S_{r-1}[r] \longrightarrow (Z_r = r)$

Broadcast attack on WPA

Motivation : Plaintext recovery

Broadcast attack

- Same plaintext encrypted using multiple random keys.
- First studied in context of RC4 by Mantin and Shamir, 2001.

Broadcast attack against RC4

- Recovery of second byte – Mantin and Shamir, 2001.
- Recovery of first 256 bytes – Maitra et al., 2011.
- Plaintext recovery attack on RC4 – Isobe et al., 2013.
- Plaintext recovery attack on TLS – AlFardan et al., 2013.
- Plaintext recovery attack on WPA – Paterson et al., 2014.

Our idea : Use the known IV

Existing approach

- Capture a number of ciphertext bytes in broadcast scenario.
- Use known biases of the form $(Z_r = v)$ to recover P_r.
- Use all known biases in keystream to improve the recovery.

Our approach

- Recall : $K[0], K[1], K[2]$ are constructed from the IV.
- IV is public; hence $K[0], K[1], K[2]$ are known in each case.

Intuition : Plaintext recovery may be improved for WPA by exploiting the knowledge of the key bytes K[0], K[1], K[2].

Exploiting knowledge of $K[0], K[1], K[2]$

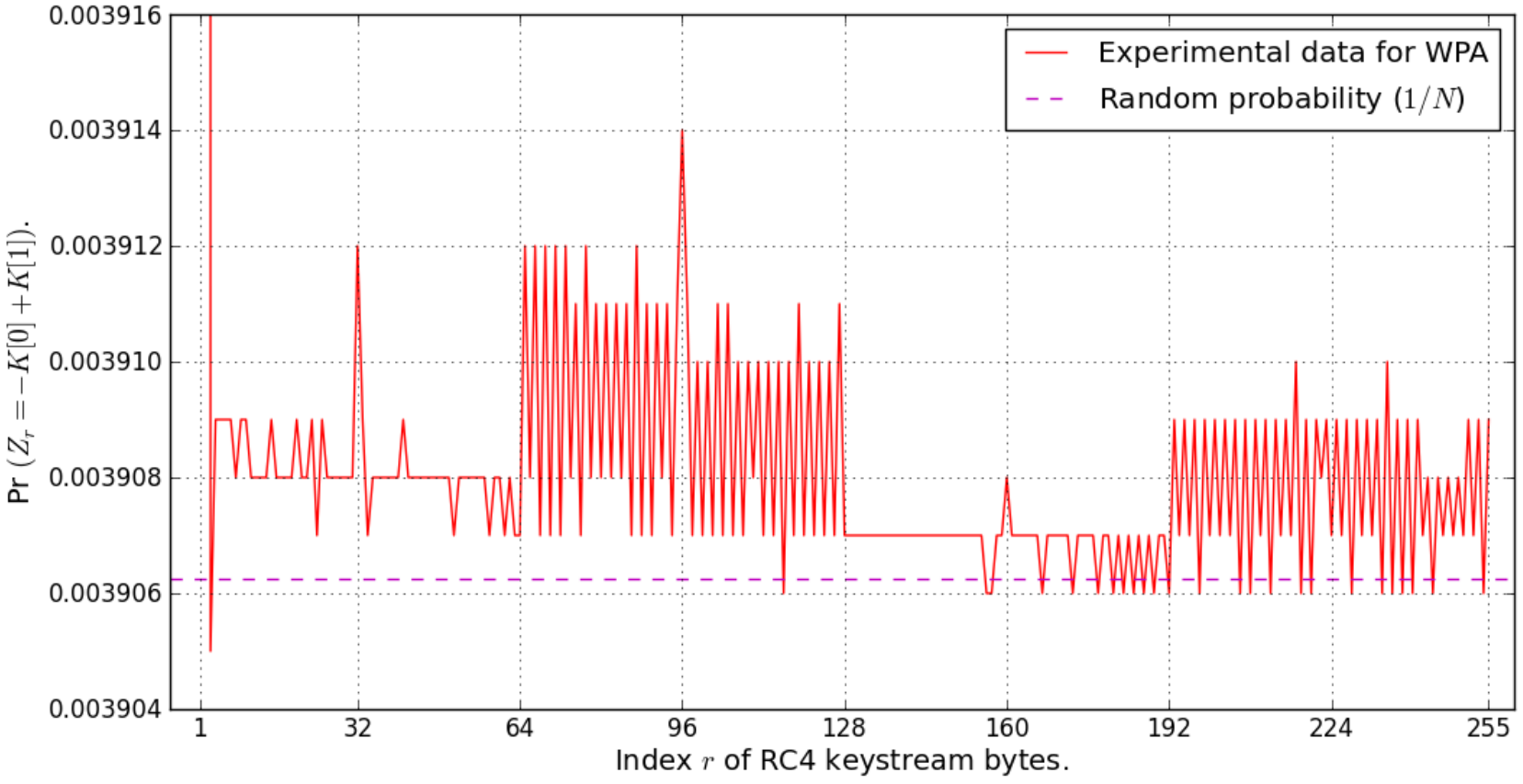
- Existing attacks use biases of keystream to absolute values.
- We explore correlations of keystream bytes with linear combinations of the known values $K[0], K[1], K[2]$.

Goal : exploit biases of following form for broadcast attack

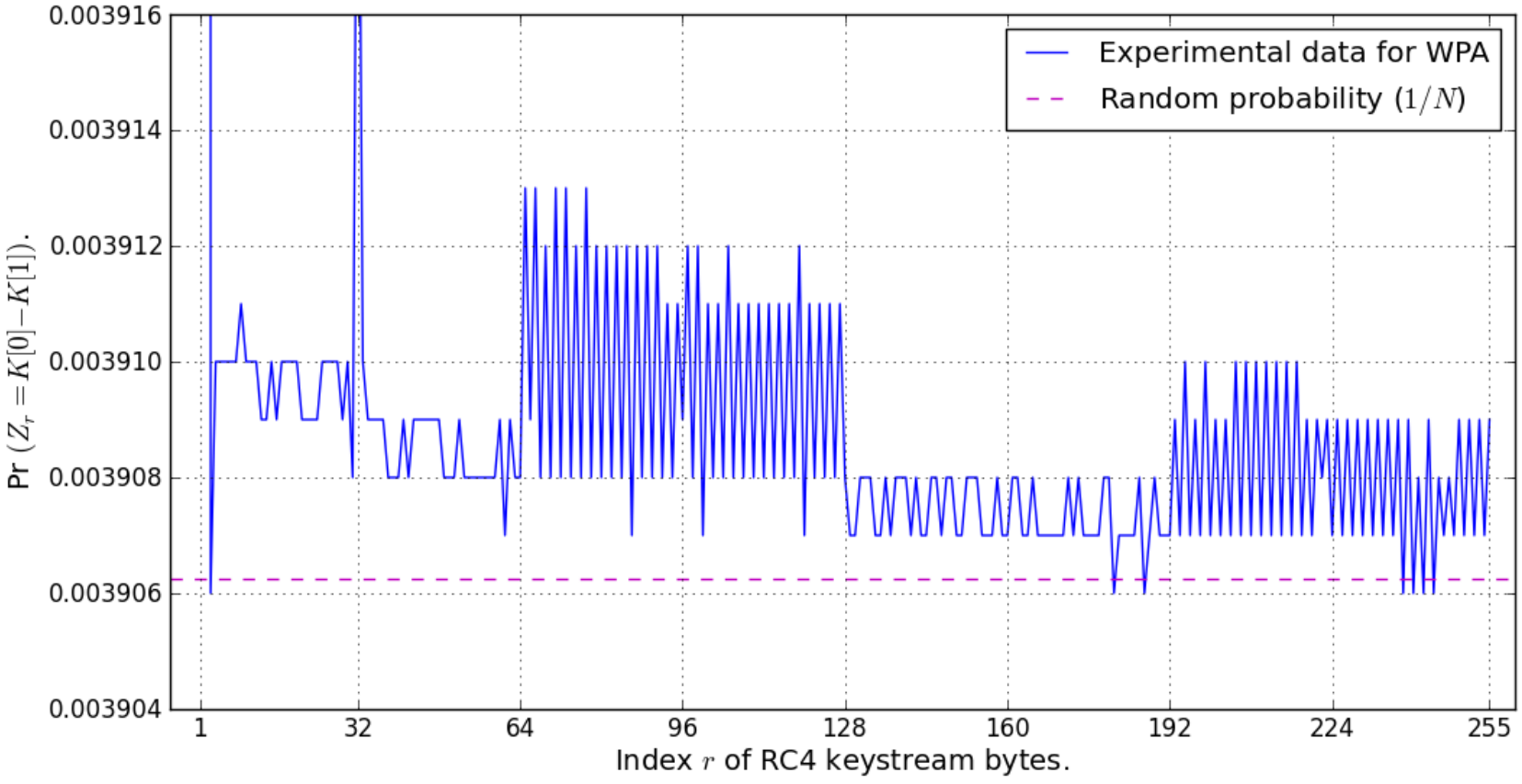
$$Z_r = a \cdot K[0] + b \cdot K[1] + c \cdot K[2] + d$$

$r \in [1, 257], \qquad a, b, c \in \{-1, 0, 1\}, \qquad d \in \{-3, -2, -1, 0, 1, 2, 3\}$

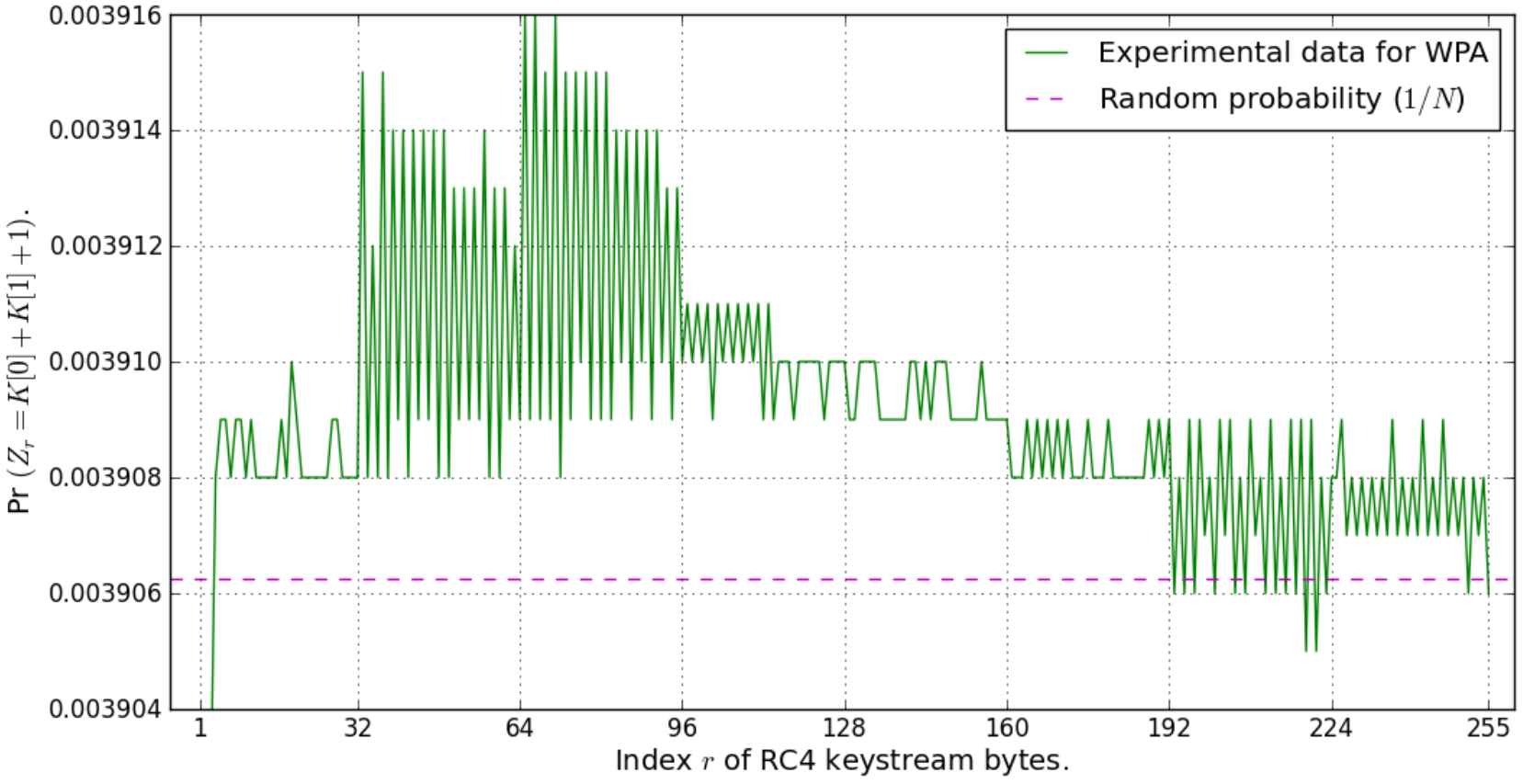
Observation : Biases in $Z_r = -K[0] + K[1]$



Observation : Biases in $Z_r = K[0] - K[1]$



Observation : Biases in $Z_r = K[0] + K[1] + 1$



Observation : Specific biases

Byte	Linear combinations	Data
Z_1	$-K[0] - K[1]$	0.005338
	$K[0]$	0.004179
	$K[0] + K[1] + K[2] + 3$	0.004633
	$K[0] + K[1] + 1$	0.003760
	$K[0] - K[1] - 1$	0.003905
	$K[2] + 3$	0.003902
	$-K[0] - K[1] + K[2] + 3$	0.003903
Z_2	$-1 - K[0] - K[1] - K[2]$	0.005303
	$-K[1] - K[2] - 3$	0.005314
	$K[1] + K[2] + 3$	0.005315
	$K[0] + K[1] + K[2] + 3$	0.002503
Z_3	$K[0] + K[1] + K[2] + 3$	0.004405
Z_{256}	$-K[0]$	0.004429
	$-K[1]$	0.004036
Z_{257}	$-K[0] - K[1]$	0.004094

Broadcast attack on WPA

Byte	Biased event	Samples
Z_1	$Z_1 = -K[0] - K[1]$, $Z_1 = K[0] + K[1] + K[2] + 3$	$5 \cdot 2^{13}$
Z_2	$Z_2 = 0$	2^{14}
Z_3	$Z_3 = K[0] + K[1] + K[2] + 3$	2^{19}
Z_{256}	$Z_{256} = -K[0]$	2^{19}
Z_{257}	$Z_{257} = -K[0] - K[1]$	2^{21}

Implication of this result

- Significant improvement in recovering bytes $\{1, 3, 256, 257\}$.
- Existing works require around 2^{30} samples for the same.

Summary of contributions

Biases in RC4

- Proof for $Z_r = r$, observed by Isobe et al., 2013.
- Observation and proof of bias in $Z_{259} = 3$.

Biases in WPA

- Proof for $Z_1 = v$, observed by AlFardan et al., 2013.
- Significantly improved WPA distinguisher with complexity 2^{19}.
- Proof for $Z_r = 0$, observed by AlFardan et al., 2013.

IV-dependence in WPA

- Correlation of keystream bytes to first three bytes of RC4 key.
- Larger biases in WPA than the known absolute biases.
- Improved plaintext recovery of some bytes in WPA.

Thank You!