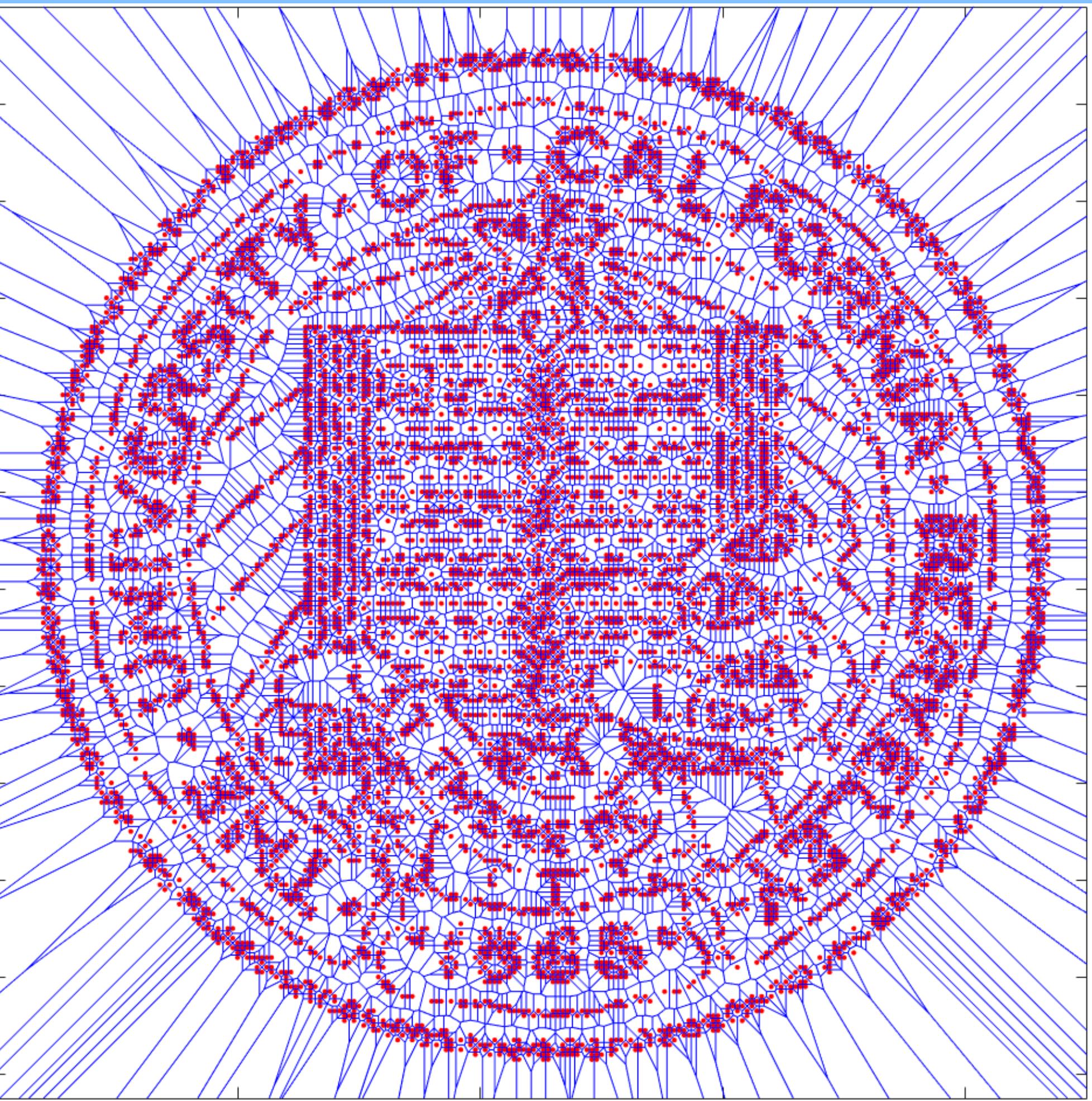


The State of the Uniform: Attacks on Encrypted Databases Beyond the Uniform Query Distribution

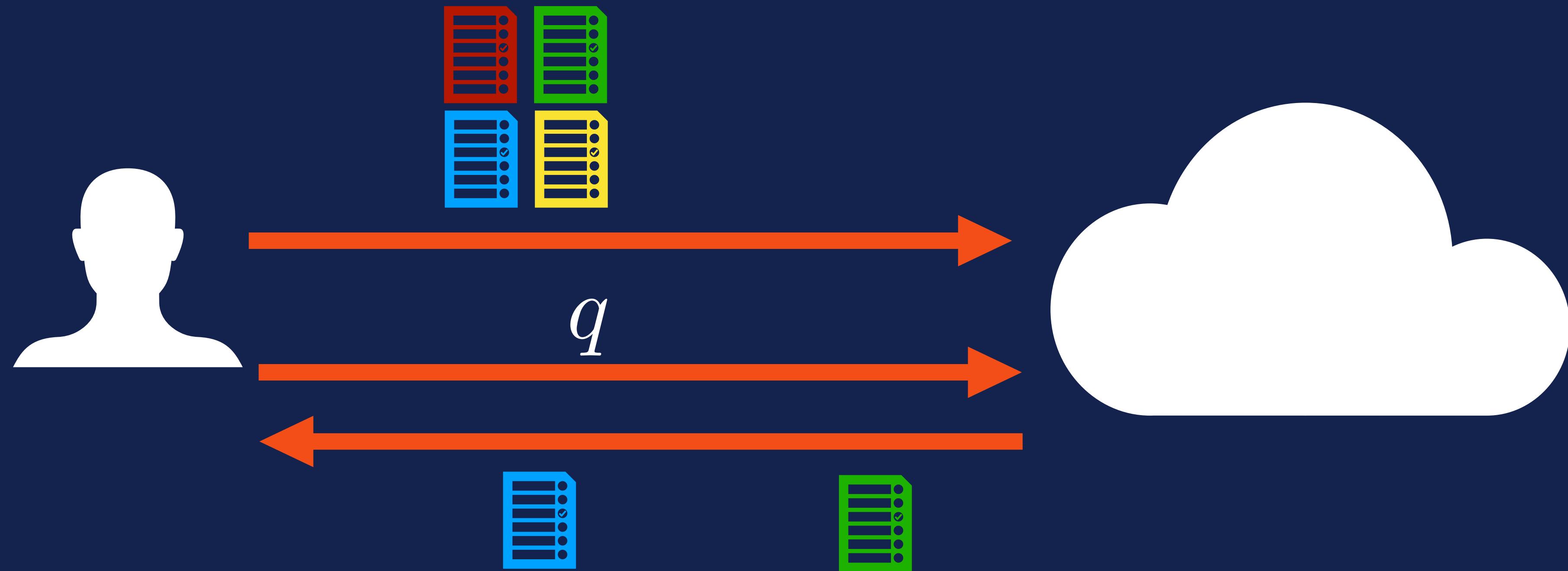
EVGENIOS M. KORNAROPOULOS
UC BERKELEY

Joint work with:
Charalampos (Babis) Papamanthou
Roberto Tamassia



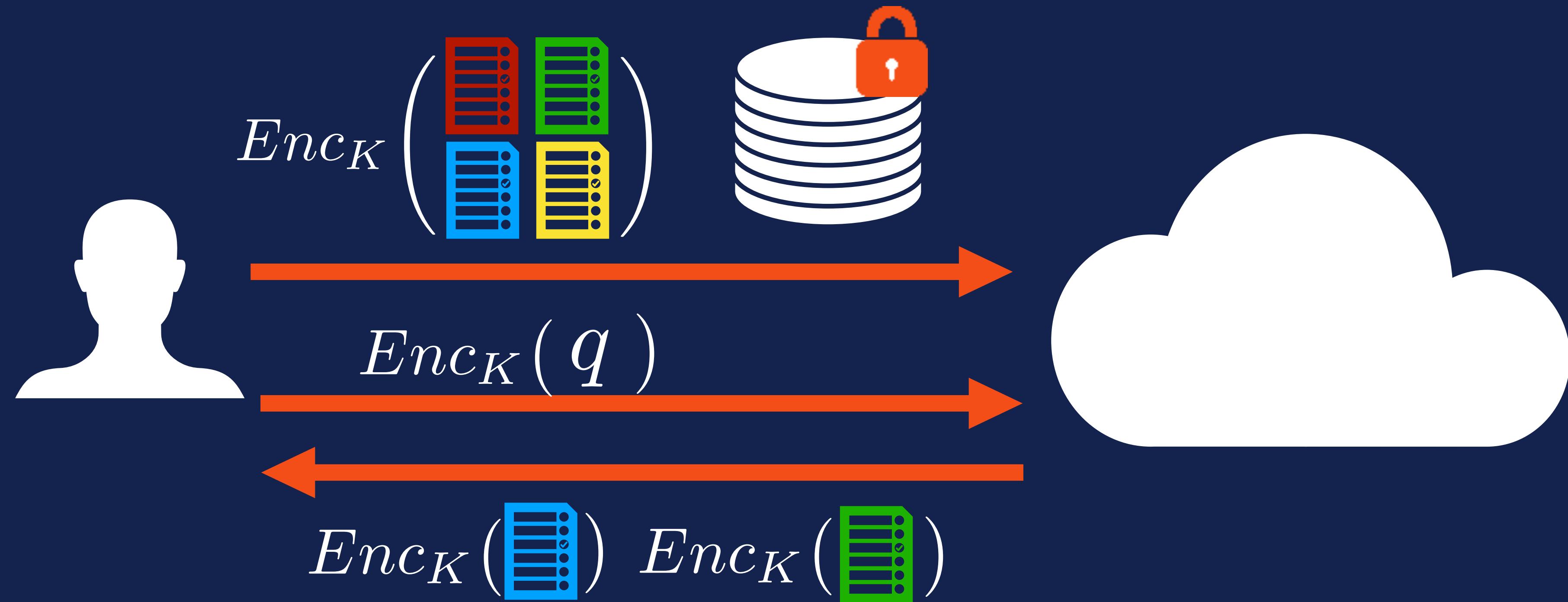


INTRO (PLAIN) SEARCH





INTRO ENCRYPTED SEARCH





INTRO HISTORY



[SWP] - S&P'00 [CGKO] - CCS'06

[BBO] - CRYPTO'07

[BCLO] - EUROCRYPT'09

[CK] - ASIACRYPT'10

[BCO] - CRYPTO'11

[KPR] - CCS'12

[IKK] - NDSS'12

[CJJKRS] - CRYPTO'13

[KP] - FC'13

[PLZ] - S&P'13

[CJJJKRS] - NDSS'14

[PKVKMCGKB] - S&P'14

[NPG] - S&P'14

[CT] - EUROCRYPT'14

[SPS] - NDSS'14

[FJKNRS] - ESORICS'15

[MCOKC] - SIGMOD'15

[MKNK] - CCS'15

[CGPR] - CCS'15

[NWKW] - CCS'15

[ANSS] - STOC'16

[KKNO] - CCS'16

[B] - CCS'16

[ZKP] - USENIX SEC'16

[DPPDG] - SIGMOD'16

[KM] - EUROCRYPT'17

[BMO] - CCS'17

[DP] - SIGMOD'17

[GLMP] - CCS'18

[CPPJ] - CCS'18

[KM] - ASIACRYPT'18

[DPP] - CRYPTO'18

[KMO] - CRYPTO'18

[ASS] - CRYPTO'18

[LMP] - S&P'18

[BT] - PETS'19

[AKM] - PETS'19

[AHKM] - PETS'19

[AHKM] - EUROCRYPT'19

[PBP] - VLDB'19

[GLMP] - S&P'19

[KPT] - S&P'19

[PPYY] - CCS'19

[GJW] - CCS'19

[PWLP] - EuroS&P'20

[BKM] - NDSS'20

[DCPP] - NDSS'20

[DDPS] - USENIX SEC'20

[KPT] - S&P'20



INTRO HISTORY



[SWP] - S&P'00 [CGKO] - CCS'06

[BBO] - CRYPTO'07

[BCLO] - EUROCRYPT'09

[CK] - ASIACRYPT'10

[BCO] - CRYPTO'11

[KPR] - CCS'12

[IKK] - NDSS'12

[CJJKRS] - CRYPTO'13

[KP] - FC'13

[PLZ] - S&P'13

[CJJJKRS] - NDSS'14

[PKVKMCGKB] - S&P'14

[NPG] - S&P'14

[CT] - EUROCRYPT'14

[SPS] - NDSS'14

[FJKNRS] - ESORICS'15

[MCOKC] - SIGMOD'15

[MKNK] - CCS'15

[CGPR] - CCS'15

[NKW] - CCS'15

[ANSS] - STOC'16

[KKNO] - CCS'16

[B] - CCS'16

[ZKP] - USENIX SEC'16

[DPPDG] - SIGMOD'16

[KM] - EUROCRYPT'17

[BMO] - CCS'17

[DP] - SIGMOD'17

[GLMP] - CCS'18

[CPPJ] - CCS'18

[KM] - ASIACRYPT'18

[DPP] - CRYPTO'18

[KMO] - CRYPTO'18

[ASS] - CRYPTO'18

[LMP] - S&P'18

[BT] - PETS'19

[AKM] - PETS'19

[AHKM] - PETS'19

[AHKM] - EUROCRYPT'19

[PBP] - VLDB'19

[GLMP] - S&P'19

[KPT] - S&P'19

[PPYY] - CCS'19

[GJW] - CCS'19

[PWLP] - EuroS&P'20

[BKM] - NDSS'20

[DCPP] - NDSS'20

[DDPS] - USENIX SEC'20

[KPT] - S&P'20



INTRO

WHAT IS LEAKAGE?

Client



Server

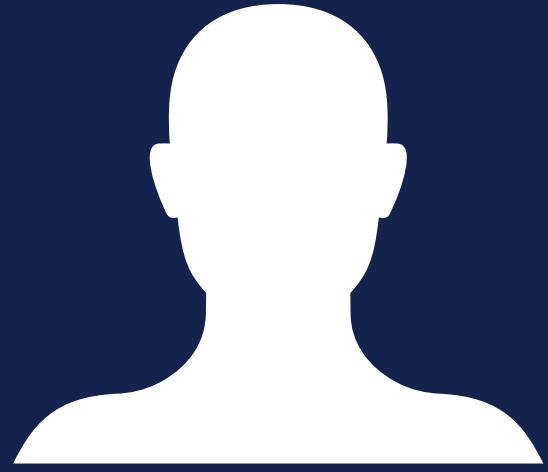




INTRO

WHAT IS LEAKAGE?

Client



Server



Tokens

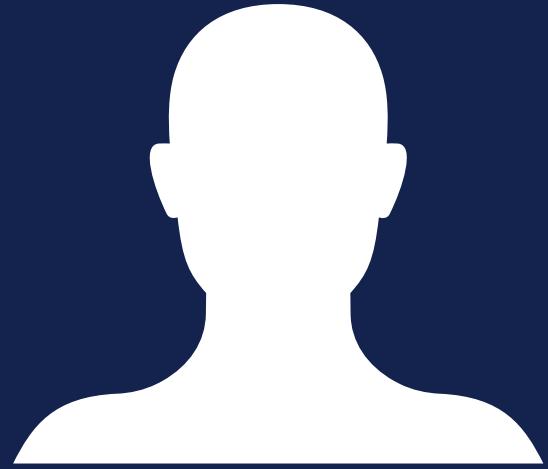
Responses



INTRO

WHAT IS LEAKAGE?

Client



Server



Tokens

$$\text{PRF}_K(\bullet) = t$$



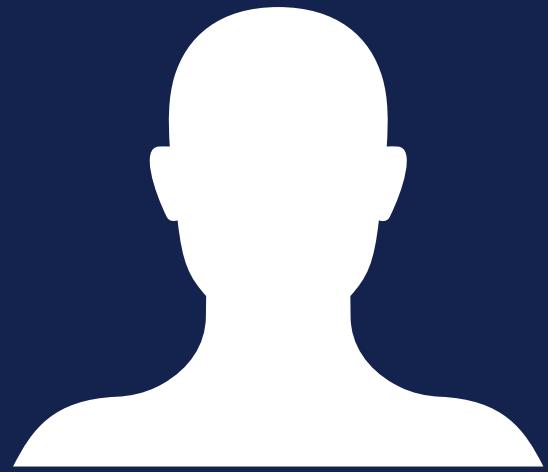
Responses



INTRO

WHAT IS LEAKAGE?

Client



Server



Tokens

$$\text{PRF}_K(\bullet) = t$$

$$\text{PRF}_K(\bullet) = t'$$

$$\text{PRF}_K(\bullet) = t''$$

$$\text{PRF}_K(\bullet) = t$$

Responses





INTRO

WHAT IS LEAKAGE?

Client



Server



Tokens

$$\text{PRF}_K(\bullet) = t$$

$$\text{PRF}_K(\bullet) = t'$$

$$\text{PRF}_K(\bullet) = t''$$

$$\text{PRF}_K(\bullet) = t$$

Responses

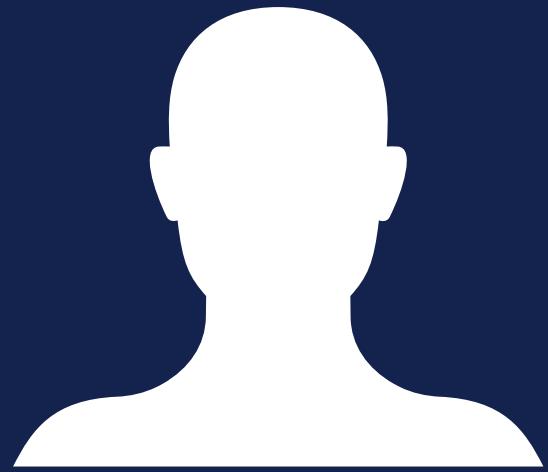




INTRO

WHAT IS LEAKAGE?

Client



Server



Tokens

$$\text{PRF}_K(\bullet) = t$$

$$\text{PRF}_K(\bullet) = t'$$

$$\text{PRF}_K(\bullet) = t''$$

$$\text{PRF}_K(\bullet) = t$$

**Search Pattern
Leakage**

Responses

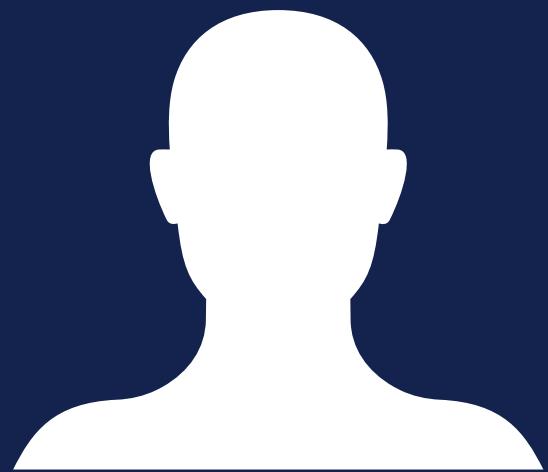




INTRO

WHAT IS LEAKAGE?

Client



Server



Tokens

$$\text{PRF}_K(\bullet) = t$$

$$\text{PRF}_K(\bullet) = t'$$

$$\text{PRF}_K(\bullet) = t''$$

$$\text{PRF}_K(\bullet) = t$$

Search Pattern
Leakage

Responses



Access Pattern Leakage



INTRO

LEAKAGE-ABUSE ATTACKS: STATE OF THE ART



INTRO

LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions					Dense Database
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution		
KPT	k -NN	Uniform	-	-	-	-	-

S&P'19

Data Recovery on Encrypted Databases With k -Nearest Neighbor Query Leakage

Evgeneios M. Kounoupis
Brown University
evgenios@cs.brown.edu

Charalambos Papamanthou
University of Maryland
cgp@umiacs.umd.edu

Roberto Tamassia
Brown University
rtt@cs.brown.edu

Abstract. Recent work by Kellam et al. (CCS'16) and Lichman et al. (ASIACRYPT'18) demonstrated attacks of data recovery for encrypted databases that support rich queries such as range queries. In this paper we discuss the first data recovery attack on encrypted databases supporting one-dimensional k -nearest neighbor (k -NN) queries, which are widely used in spatial data management. Our attack exploits a generic k -NN query leakage primitive that attacker observes the identities of matched records. We consider both interleaved records, where the leakage is a set, and ordered records, where the leakage is a k -tuple ordered by distance from the query point.

As a first step, we perform a theoretical feasibility study on exact reconstruction, i.e., recovery of the exact plaintext values of the encrypted database. For ordered responses, we show that exact reconstruction is feasible if the attacker has additional access to some auxiliary information that is normally not available in encrypted databases. For ordered responses, we prove that exact reconstruction is feasible for the interleaved case of valid reconstruction. As a next step, we propose practical and more realistic approximate reconstruction attacks as to recover an approximation of the claimed values. For ordered responses, we show that the sum leakage,



INTRO LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions					Dense Database
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution		
KPT	k -NN	Uniform	-	-	-	-	-
KKNO	Range	Uniform	-	-	-	-	-

CCS'16

Generic Attacks on Secure Outsourced Databases

George Kellaris¹
Boston University and
Harvard University
gkellari@cs.harvard.edu

George Kellaris¹
Boston University
gkellari@cs.harvard.edu

Kobi K Nissim²
Bar-Gurion University and Harvard University
kobik@seas.harvard.edu

Adam O'Neill
Georgetown University
adam@cs.georgetown.edu

ABSTRACT

Recently, various protocols have been proposed for securely outsourcing databases to a third party server, ranging from systems with TCB-hybrid security based on strong cryptographic primitives such as fully homomorphic encryption or oblivious RAM, to more practical implementations based on searchable symmetric encryption or even on deniable and zero-knowledge encryption. On the flip side, various attacks have emerged that show that for many of these protocols confidentiality of the data can be compromised while giving certain auxiliary information.

We take a step back and identify a need for a formal analysis of leakage and abuse of the data itself in outsourced database systems, independent of the details of the system. We propose shorthash methods that capture all the information stored in sufficient generality and identify two low-cost sources of leakage: namely access patterns

INTRODUCTION

As organizations struggle with the accumulation of large amounts of data, a popular practice is to outsource them to third party servers. Because their data may be sensitive or valuable in nature, most outsourced responses we give that can be used to extract information about the data itself are valid reconstruction. As a result, we propose practical and more realistic operational reconstruction attacks as to recover an amalgamation of the claimed values. For ordered data stores, such as the sum leakage.

Data Recovery on Encrypted Databases With k -Nearest Neighbor Query Leakage

Eugenios M. Kranakisopoulos
Brown University
eugenios@cs.brown.edu

Charalampus Papamanthou
University of Maryland
cpcp@umiacs.umd.edu

Roberto Tamassia
Brown University
rtt@cs.brown.edu

[46], demonstrate how an attacker can utilize access patterns to launch query-recovery attacks under various assumptions.

However, in the case of richer queries (e.g., range [16], [23], [37] and SQL [34], [38]), more severe data-recovery attacks are possible due to the expressiveness of the query. In particular, the work by Kellaris, Kollaris, Nissim, and O'Neill [25] attacks SIMD-type systems that support range queries (e.g., [16], [21], [29]) by observing record identifiers whose plaintext values belong to the queried range. Similarly, a recent work by Lacharité, Minard, and Paterson [27] further explores range query leakage to retrieve exact and approximate reconstruction in the case of dense datasets with orders of magnitude fewer queries (when compared to [25]). Finally, order-preserving encryption based systems (e.g., CryptDB [18]) supporting even more expressive queries (such as SQL) have been shown to be vulnerable to data-recovery attacks [16], [21], [23] even without observing any queries, and by the sum leakage.

As a first step, we perform a theoretical feasibility study on query reconstruction, i.e., recovery of the exact plaintext values of the encrypted database. For ordered responses, we show that exact reconstruction is feasible if the attacker has additional access to some auxiliary information that is normally not available in outsourced database systems. We prove that exact reconstruction is feasible for the following classes of valid reconstruction. As a result, we propose practical and more realistic operational reconstruction attacks as to recover an amalgamation of the claimed values. For ordered data stores, such as the sum leakage.



INTRO LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions					Dense Database
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution		
KPT	k -NN	Uniform	-	-	-	-	-
KKNO	Range	Uniform	-	-	-	-	-
LMP	Range	Agnostic	-	-	-	-	●

CCS'16

Generic Attacks on Secure Outsourced Databases

Georgia Kellaris¹
Boston University and Harvard University
gkellaris@g.harvard.edu

George Kolios²
Boston University
gkolios@seas.bu.edu

Kopki Nassim³
Bent-Gurion University and Harvard University
kcb@seas.harvard.edu

Adam O'Neill
Georgetown University
adamo@cs.georgetown.edu

ABSTRACT
Recently, various protocols have been proposed for securely outsourcing database storage to a third party server, ranging from systems with TEE-based security based on strong cryptographic primitives such as fully homomorphic encryption or oblivious RAM, to more practical implementations based on searchable symmetric encryption or even on deterministic and entropy-preserving encryption. On the flip side, various attacks have emerged that show that for some of these protocols confidentiality of the data can be compromised while giving certain auxiliary information.

We take a step back and identify a need for a formal analysis of leakage and abuse of auxiliary information itself in outsourced database systems, independent of the details of the systems. We propose sharing methods that capture all the information stored in sufficient generality, and illustrate two lower bounds of leakage, namely access patterns

S&P'18

Improved Reconstruction Attacks on Encrypted Data Using Range Query Leakage

2018 IEEE Symposium on Security and Privacy
Marie-Sarah Lachaud, Bruce Minace, Kenneth G. Paterson
Information Security Group
Royal Holloway, University of London
Egham, United Kingdom
Email: [marie-sarah.lachaud, bruce.minace, kenneth.paterson]@rhul.ac.uk

ABSTRACT
Recent work by Kellaris et al. (CCS'16) and Lachaud et al. (S&P'18) demonstrated attacks of data recovery for encrypted databases that support rich queries such as range queries. In this paper we discuss the first data recovery attack on encrypted databases supporting one-dimensional k-nearest neighbor (k -NN) queries, which are widely used in spatial data management. Our attack exploit a generic k -NN query leakage primitive that allows observers the identities of matched records. We consider both insertion attacks, where the leakage is a set, and removal attacks, where the leakage is a tuple ordered by distance from the query point.

As a first step, we perform a theoretical feasibility study on range reconstruction, i.e., recovery of the exact plaintext values of the encrypted database. For ordered responses, we show that exact reconstruction is *feasible* if the attacker has additional access to some auxiliary information that is normally not available in encrypted databases. Range reconstruction is possible within an expected number of queries $\mathcal{O}(N \log N + O(N))$, where N is the number of distinct identifiers rather than legacy indices, promise to do better, in the sense of providing lessening less overhead for an amortization of the claimed values. For ordered

S&P'19

Data Recovery on Encrypted Databases With k -Nearest Neighbor Query Leakage

Eugenios M. Karampoulos
Brown University
ekarampoulos@cs.brown.edu

Charalampus Papamanthou
University of Maryland
cpap@umiacs.umd.edu

Roberto Tamassia
Brown University
rt@cs.brown.edu

ABSTRACT
Recent work by Kellaris et al. (CCS'16) and Lachaud et al. (S&P'18) demonstrated attacks of data recovery for encrypted databases that support rich queries such as range queries. In this paper we discuss the first data recovery attack on encrypted databases supporting one-dimensional k-nearest neighbor (k -NN) queries, which are widely used in spatial data management. Our attack exploit a generic k -NN query leakage primitive that allows observers the identities of matched records. We consider both insertion attacks, where the leakage is a set, and removal attacks, where the leakage is a tuple ordered by distance from the query point.

However, in the case of richer queries (e.g., range [16], [22], [27] and SQL [36], [38]), more severe data-recovery attacks are possible due to the expressiveness of the query. In particular, the work by Kellaris, Kolios, Nassim, and O'Neill [25] attacks SIMD-type systems that support range queries (e.g., [16], [21], [29]) by observing record identifiers whose plaintext values belong to the queried range. Similarly, a recent work by Lachaud, Minace, and Paterson [27] further explores range query leakage to retrieve range and approximate reconstruction in the case of dense datasets with *orders of magnitude fewer queries* (when compared to [25]). Finally, order-preserving encryption based systems (e.g., CryptDB [18]) supporting even more expressive queries (such as SQL) have been shown to be vulnerable to data-recovery attacks [16], [22], [23] even without observing any queries, and in the same fashion.



INTRO LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions					Dense Database
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution		
KPT	k -NN	Uniform	-	-	-	-	-
KKNO	Range	Uniform	-	-	-	-	-
LMP	Range	Agnostic	-	-	-	-	●
GLMP GENERALIZED KKNO	Range	Uniform	-	-	-	-	-
GLMP APPROX VALUE	Range	Uniform	●	-	-	-	-
GLMP AOR to ADR	Range	Agnostic	-	●	●	●	-

CCS'16

Generic Attacks on Secure Outsourced Databases

Georgia Kellaris¹
Boston University and Harvard University
gkellari@cs.harvard.edu

George Kellaris¹
Boston University
gkellari@cs.harvard.edu

Kopki Nasim²
Bent-Gurion University and Harvard University
kcbb@seas.harvard.edu

Adam O'Neill³
Georgetown University
adam@cs.georgetown.edu

ABSTRACT

Recently, various protocols have been proposed for securely outsourcing database storage to a third party server, ranging from systems with fully-homomorphic encryption to semi-homomorphic encryption or oblivious RAM. In this initial work, we provide experimental results demonstrating the efficacy of our attack on real datasets with a variety of different schemes. On all these datasets, after the required number of queries our attack successfully reconstructs the secret attributes of every record in at most a few seconds.

KEYWORD

generic

attacks

outsourced

databases

leakage

abuse

attacks

security

S&P'18

Improved Reconstruction Attacks on Encrypted Data Using Range Query Leakage

Marie-Sarah Lachaud, Brice Minard, Kenneth G. Paterson
Information Security Group
Royal Holloway, University of London
Egham, United Kingdom
Email: {marie-sarah.lachaud, brice.minard, kenneth.g.paterson}@rhul.ac.uk

ABSTRACT

We analyse the security of database encryption schemes supporting range queries against practical adversaries. The bulk of our work applies in a prove setting, where the adversary's view is limited to records matched by each query (known as active pattern leakage). We also consider a more specific setting where raw information is also leaked, which is inherent to multiple recent encryption schemes supporting range queries. We provide three attacks.

First, we consider an adversary which aims to reconstruct database values using a linear search. For example, deterministic encryption allows matching queries to be made, while Order-Preserving Encryption (OPE) allow range queries to be efficiently supported.

At the same time, our understanding of the security of such schemes often against various kinds of adversary is still developing. This has led to various attacks being found against some of the early schemes [11, 12, 13, 14, 15, 16, 17] – a good summary of this line of research is available in [8]. A second generation of schemes, which typically use custom indices rather than legacy indices, promise to do better, in the sense of providing less leakage less

support efficient search. For example, deterministic encryption allows matching queries to be made, while Order-Preserving Encryption (OPE) allow range queries to be efficiently supported.

At the same time, our understanding of the security of such schemes often against various kinds of adversary is still developing. This has led to various attacks being found against some of the early schemes [11, 12, 13, 14, 15, 16, 17] – a good summary of this line of research is available in [8]. A second generation of schemes, which typically use custom indices rather than legacy indices, promise to do better, in the sense of providing less leakage less

S&P'19

Learning to Reconstruct: Statistical Learning Theory and Encrypted Database Attacks

Paul Grubbs*, Marie-Sarah Lachaud, Brice Minard, Kenneth G. Paterson
Cornell University, pg225@cornell.edu
Royal Holloway, University of London, marie-sarah.lachaud.51.5, kenneth.g.paterson@rhul.ac.uk
*From: NIST Special Publication 800-53, BSI, University of Erlangen-Nürnberg, Erlangen, Germany

ABSTRACT

We show that the problem of reconstructing encrypted databases from active pattern leakage is closely related to statistical learning theory. This new viewpoint enables us to develop broader attacks that are inspired by theoretical performance analyses. As an introduction to this viewpoint, we first present a general reduction from reconstruction with known active patterns to learning with one-dimensional loss functions. This reduction is based on a generalization of the problem of separating data points (so-called ℓ_0 -norm) from many query leakage instances where query leakage only with the relative error, and its interpretation of the size of the database, N , the number M of possible values of plain items. This reduction can be applied to any scheme that supports range queries. We show that for some datasets, full reconstruction is possible within an expected number of queries $\mathcal{O}(N \log N + O(N))$, where N is the number of distinct plaintext values. This directly improves on a quadratically bound bound in the same setting by Kellaris et al.

ABSTRACT

If an encrypted database supports a certain class of queries, but lacks the access pattern, then how damaging is that leakage as a function of the query and data distribution and number of queries?

S&P'19

Data Recovery on Encrypted Databases With k -Nearest Neighbor Query Leakage

Eugenios M. Karampoulos
Brown University
eugenios@cs.brown.edu

Charalampis Papamanthou
University of Maryland
cp@cs.umd.edu

Roberto Tamassia
Brown University
rtt@cs.brown.edu

ABSTRACT Recent works by Kellaris et al. (CCS'16) and Lachaud et al. (S&P'18) demonstrated attacks of data recovery for encrypted databases that support range queries such as range queries. In this paper we discuss the first data recovery attacks for encrypted databases that support k -nearest neighbor (kNN) queries, which are widely used in spatial data management. Our attacks exploit a generic kNN query leakage primitive that allows recovering the identities of matched records. We consider both insertion attacks, where the leakage is a set, and removal attacks, where the leakage is a tuple ordered by distance from the query point.

As a first step, we perform a theoretical feasibility study on range reconstruction, i.e., recovery of the exact plaintext values of the encrypted database. For ordered responses, we show that exact reconstruction is feasible if the attacker has additional access to some auxiliary information that is normally not available in practice. For unsorted responses, we prove that exact reconstruction is feasible if the auxiliary information (such as SVD) have been shown to be vulnerable to data-recovery attacks [16, 21, 23] even without observing enough range queries. To achieve range and approximate reconstruction in the case of dense datasets with orders of magnitude fewer queries [when compared to [5]], finally, order-preserving encryption based systems (e.g., CryptDB [18]) supporting even more expressive queries than kNN have been shown to be vulnerable to data-recovery attacks [16, 21, 23] even without observing enough range queries. In the sum total,



INTRO

LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions				
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution	Dense Database
KPT	k -NN	Uniform	-	-	-	-
KKNO	Range	Uniform	-	-	-	-
LMP	Range	Agnostic	-	-	-	●
GLMP GENERALIZEDKKNO	Range	Uniform	-	-	-	-
GLMP APPROXVALUE	Range	Uniform	●	-	-	-
GLMP AOR to ADR	Range	Agnostic	-	●	●	-



INTRO

LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions					Dense Database
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution		
KPT	k -NN	Uniform	-	-	-	-	-
KKNO	Range	Uniform	-	-	-	-	-
LMP	Range	Agnostic	-	-	-	-	●
GLMP GENERALIZEDKKNO	Range	Uniform	-	-	-	-	-
GLMP APPROXVALUE	Range	Uniform	●	-	-	-	-
GLMP AOR to ADR	Range	Agnostic	-	●	●	-	-

[KM] - EUROCRYPT'19

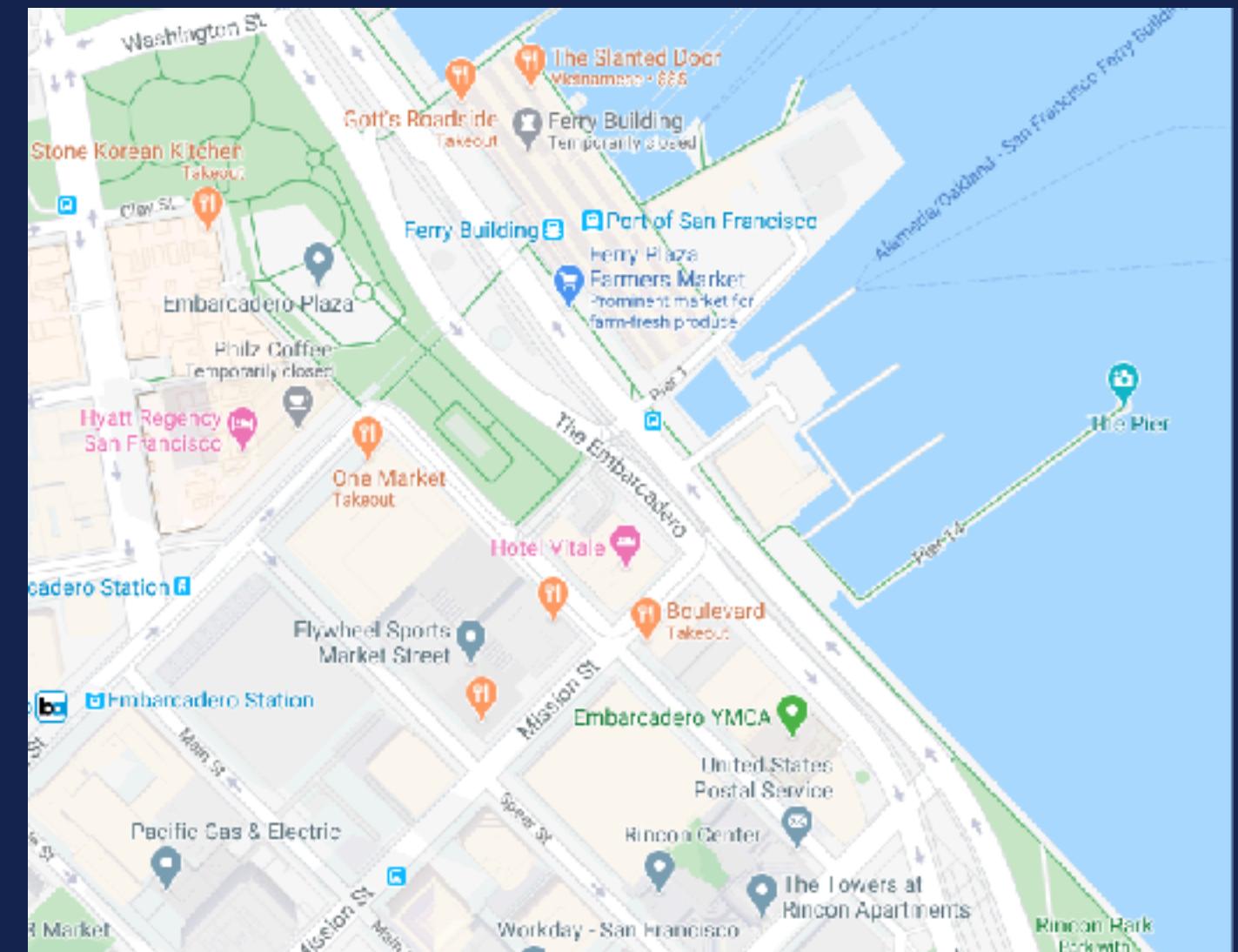
“ While there has been some progress on designing leakage attacks against STE [9, 24, 30, 32], these attacks remain mostly of theoretical interest due to the strong assumptions they rely on.”



INTRO

LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions				
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution	Dense Database
KPT	k -NN	Uniform	-	-	-	-
KKNO	Range	Uniform	-	-	-	-
LMP	Range	Agnostic	-	-	-	●
GLMP GENERALIZEDKKNO	Range	Uniform	-	-	-	-
GLMP APPROXVALUE	Range	Uniform	●	-	-	-
GLMP AOR to ADR	Range	Agnostic	-	●	●	-

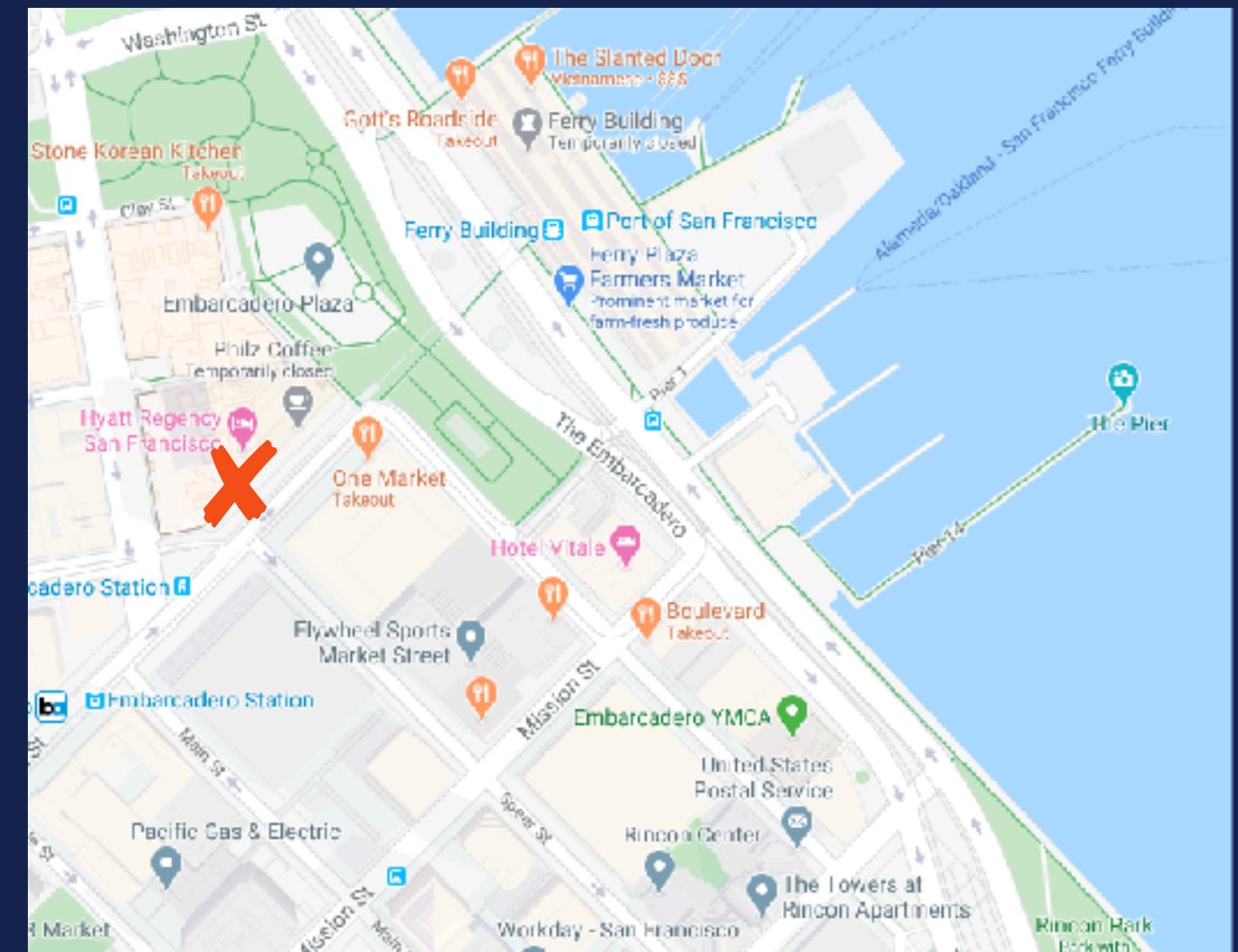




INTRO

LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions				
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution	Dense Database
KPT	k -NN	Uniform	-	-	-	-
KKNO	Range	Uniform	-	-	-	-
LMP	Range	Agnostic	-	-	-	●
GLMP GENERALIZEDKKNO	Range	Uniform	-	-	-	-
GLMP APPROXVALUE	Range	Uniform	●	-	-	-
GLMP AOR to ADR	Range	Agnostic	-	●	●	-

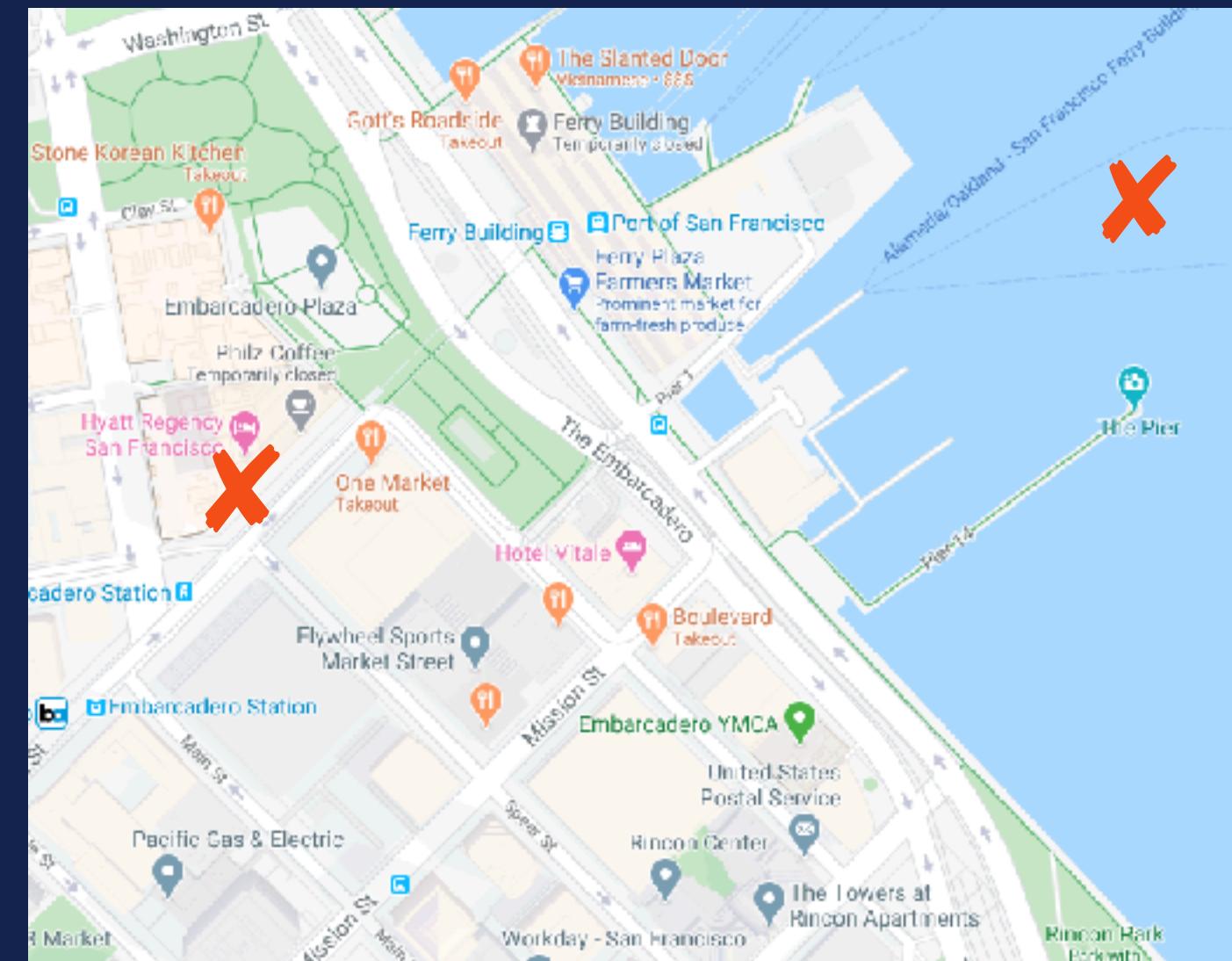




INTRO

LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions				
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution	Dense Database
KPT	k -NN	Uniform	-	-	-	-
KKNO	Range	Uniform	-	-	-	-
LMP	Range	Agnostic	-	-	-	●
GLMP GENERALIZEDKKNO	Range	Uniform	-	-	-	-
GLMP APPROXVALUE	Range	Uniform	●	-	-	-
GLMP AOR to ADR	Range	Agnostic	-	●	●	-





INTRO LEAKAGE-ABUSE ATTACKS: STATE OF THE ART

Value Reconstruction Attack Algorithms	Query Type	Assumptions					Dense Database
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution		
KPT	k -NN	Uniform	-	-	-	-	-
KKNO	Range	Uniform	-	-	-	-	-
LMP	Range	Agnostic	-	-	-	-	●
GLMP GENERALIZEDKKNO	Range	Uniform	-	-	-	-	-
GLMP APPROXVALUE	Range	Uniform	●	-	-	-	-
GLMP AOR to ADR	Range	Agnostic	-	●	●	-	-
This Work	k -NN & Range	Agnostic	-	-	-	-	-

S&P'20

The State of the Uniform: Attacks on Encrypted Databases Beyond the Uniform Query Distribution

Evgenios M. Kornaropoulos
UC Berkeley Charalampos Papamanthou
University of Maryland Roberto Tamassia
Brown University

Abstract—Recent foundational work on leakage-abuse attacks on encrypted databases has broadened our understanding of what an adversary can accomplish with a standard leakage profile. Nevertheless, all known value reconstruction attacks succeed under strong assumptions that may not hold in the real world. The most prevalent assumption is that queries are issued uniformly at random by the client. We present the first value reconstruction attacks that succeed *without any knowledge about the query or data distribution*. Our approach uses the search-pattern leakage, which exists in all known structured encryption schemes but has not been fully exploited so far. At the core of our method lies a support size estimator, a technique that utilizes the repetition of search tokens with the same response to estimate distances between encrypted values without any assumptions about the underlying distribution. We develop distribution-agnostic reconstruction attacks for both range queries and k -nearest-neighbor (k -NN) queries based on information extracted from the search-pattern leakage. Our new range attack follows a different algorithmic approach than state-of-the-art attacks, which are fine-tuned to succeed under the uniformly distributed queries. Instead, we reconstruct plaintext values under a variety of skewed query distributions and even outperform the accuracy of previous approaches under the uniform query distribution.

Fig. 1. Visual comparison between plaintext values of real world private geolocation dataset Epitix (in red) and values reconstructed by our attack AGNOSTIC-RECONSTRUCTION-KNN on k -NN queries under a Gaussian distribution and $k = 10$ (in black). Our attack achieves an approximate reconstruction (1) under a non-uniform query distribution and (2) with *any* query distribution and larger k values compared to previous work [33].

and/or data distribution. In this paper, we take the next step and demonstrate the first efficient reconstruction attacks for



STATE OF THE UNIFORM OVERVIEW

NEW INSIGHTS ON LEAKAGE EXPLOITATION

Synergy between Search Pattern Leakage + Access Pattern Leakage

Non-parametric estimation techniques on the Search Pattern Leakage information



STATE OF THE UNIFORM OVERVIEW

NEW INSIGHTS ON LEAKAGE EXPLOITATION

Synergy between Search Pattern Leakage + Access Pattern Leakage

Non-parametric estimation techniques on the Search Pattern Leakage information

REVISIT RANGE APPROXIMATE RECONSTRUCTION

Combination of new tools and Optimization formulation and no assumptions about the query or data distribution



STATE OF THE UNIFORM OVERVIEW

NEW INSIGHTS ON LEAKAGE EXPLOITATION

Synergy between Search Pattern Leakage + Access Pattern Leakage

Non-parametric estimation techniques on the Search Pattern Leakage information

REVISIT RANGE APPROXIMATE RECONSTRUCTION

Combination of new tools and Optimization formulation and no assumptions about the query or data distribution

REVISIT k-NN APPROXIMATE RECONSTRUCTION

Smaller number of samples, larger k values and no assumptions about the query or data distribution



STATE OF THE UNIFORM OVERVIEW

NEW INSIGHTS ON LEAKAGE EXPLOITATION

Synergy between Search Pattern Leakage + Access Pattern Leakage

Non-parametric estimation techniques on the Search Pattern Leakage information

REVISIT RANGE APPROXIMATE RECONSTRUCTION

Combination of new tools and Optimization formulation and no assumptions about the query or data distribution

REVISIT k-NN APPROXIMATE RECONSTRUCTION

Smaller number of samples, larger k values and no assumptions about the query or data distribution



STATE OF THE UNIFORM ASSUMPTIONS OF THE ATTACKS

BOUNDARIES:

Known boundaries α and β

STATIC:

No updates in the database

EXACT RESPONSES:

No false positives records or missing records

QUERY DISTRIBUTION:

Fixed distribution with non-zero probabilities. Queries are i.i.d.

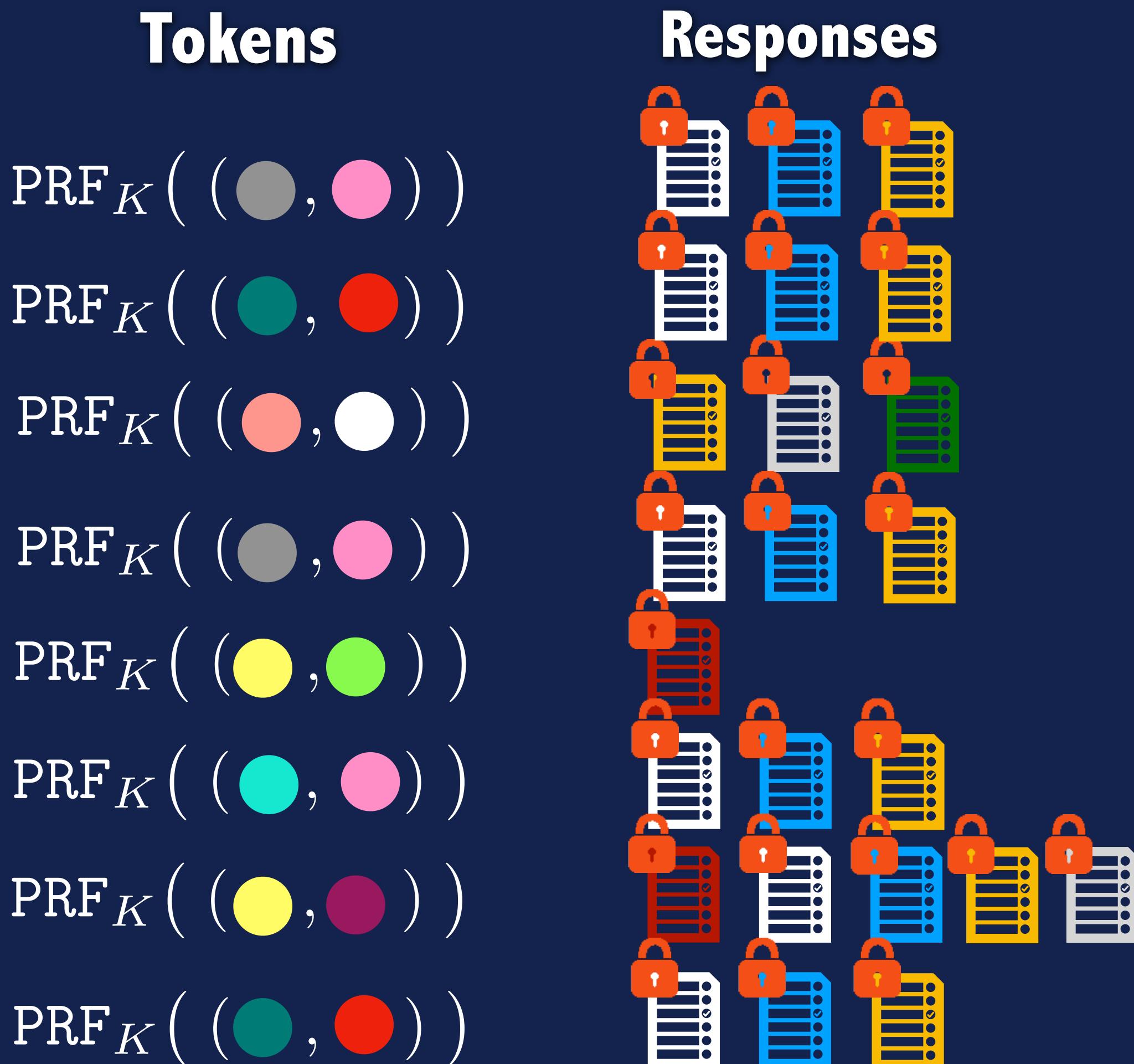


WHAT CAN THE ADVERSARY LEARN FROM THE SEARCH PATTERN LEAKAGE ?



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”

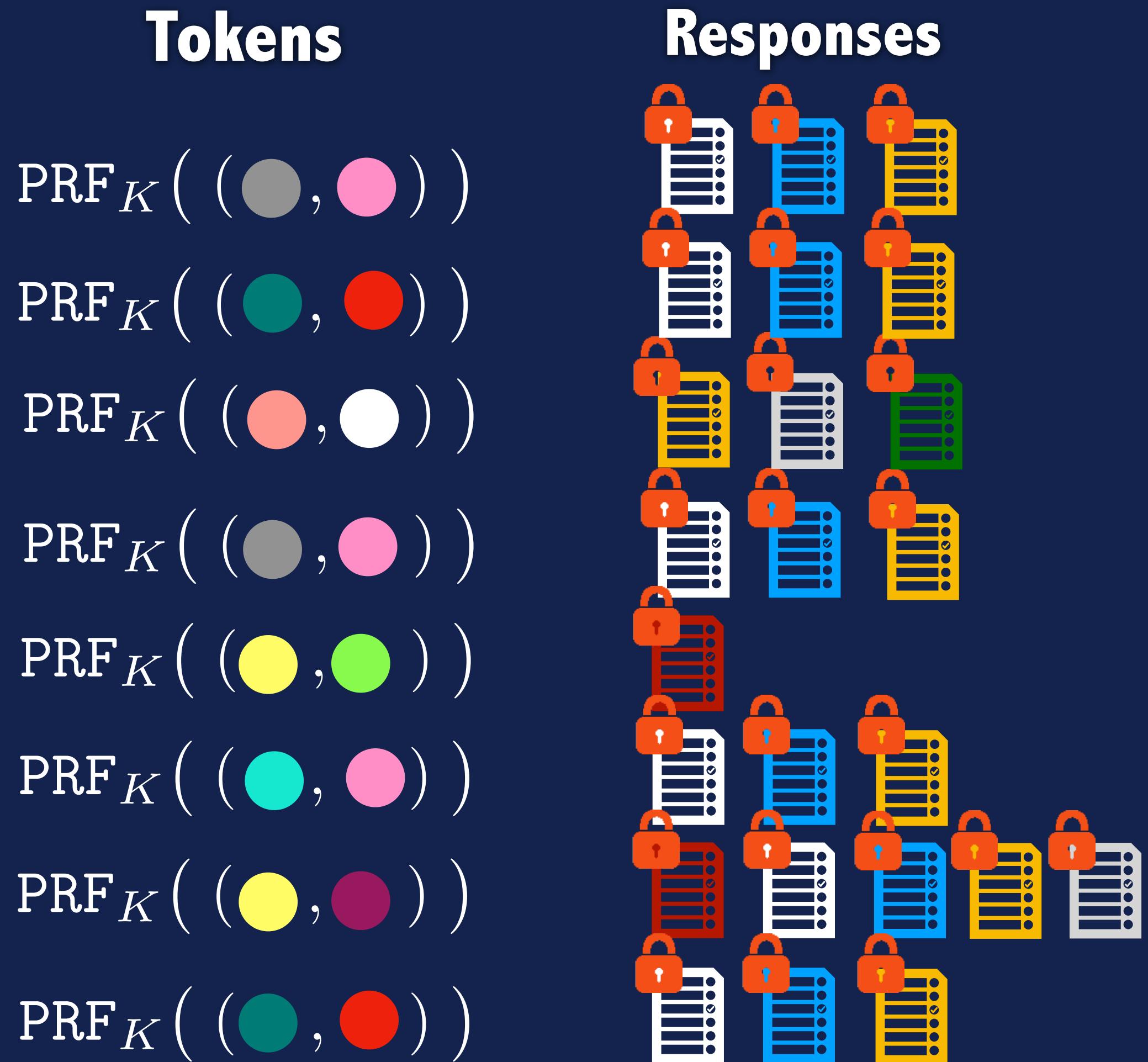




STATE OF THE UNIFORM

SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”

$\text{PRF}_K((\bullet, \circ))$

$\text{PRF}_K((\bullet, \bullet))$

$\text{PRF}_K((\circ, \circ))$

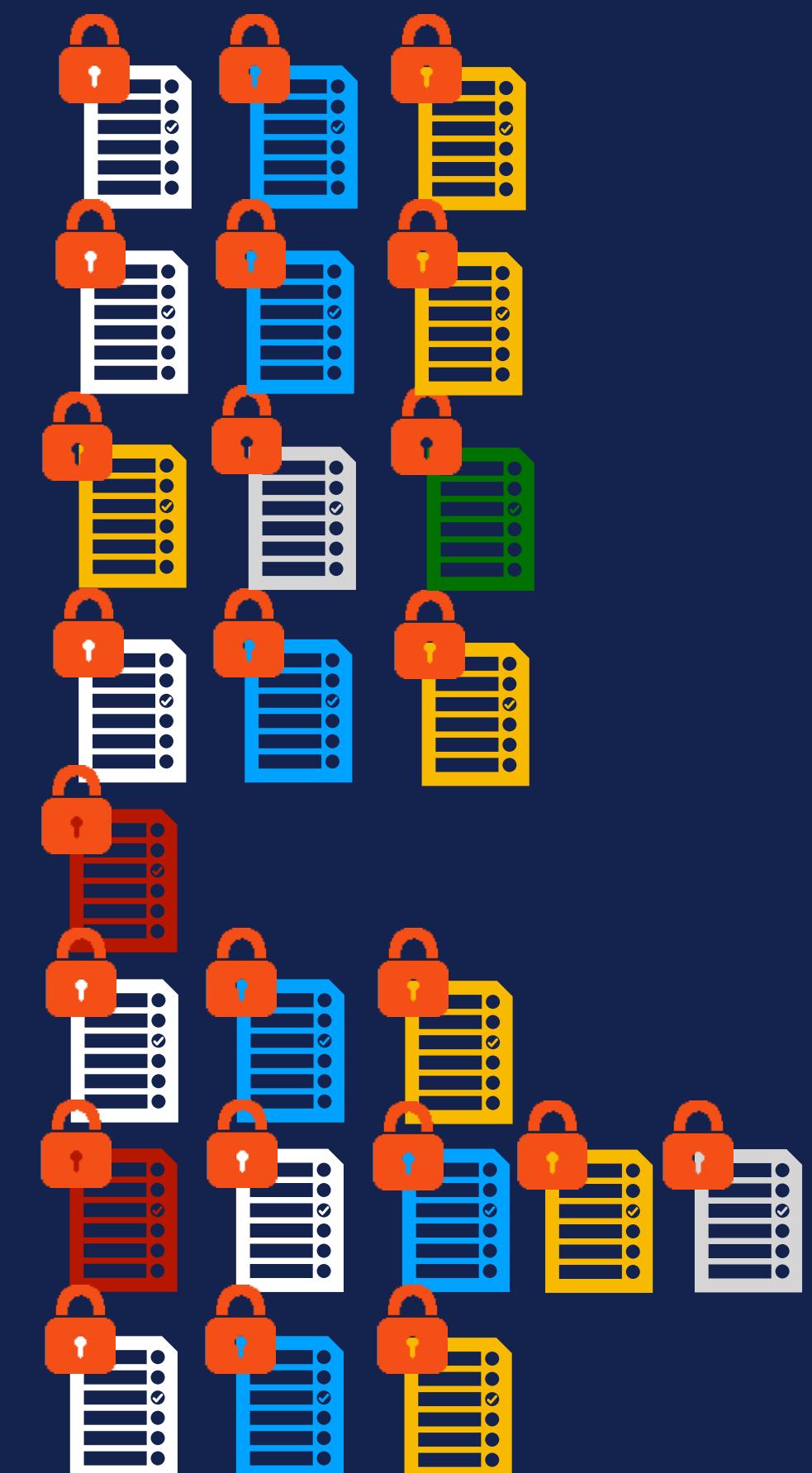
$\text{PRF}_K((\bullet, \circ))$

$\text{PRF}_K((\bullet, \bullet))$

$\text{PRF}_K((\circ, \bullet))$

$\text{PRF}_K((\bullet, \circ))$

$\text{PRF}_K((\bullet, \bullet))$





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response

$\text{PRF}_K((\bullet, \circ))$

$\text{PRF}_K((\bullet, \bullet))$

$\text{PRF}_K((\circ, \circ))$

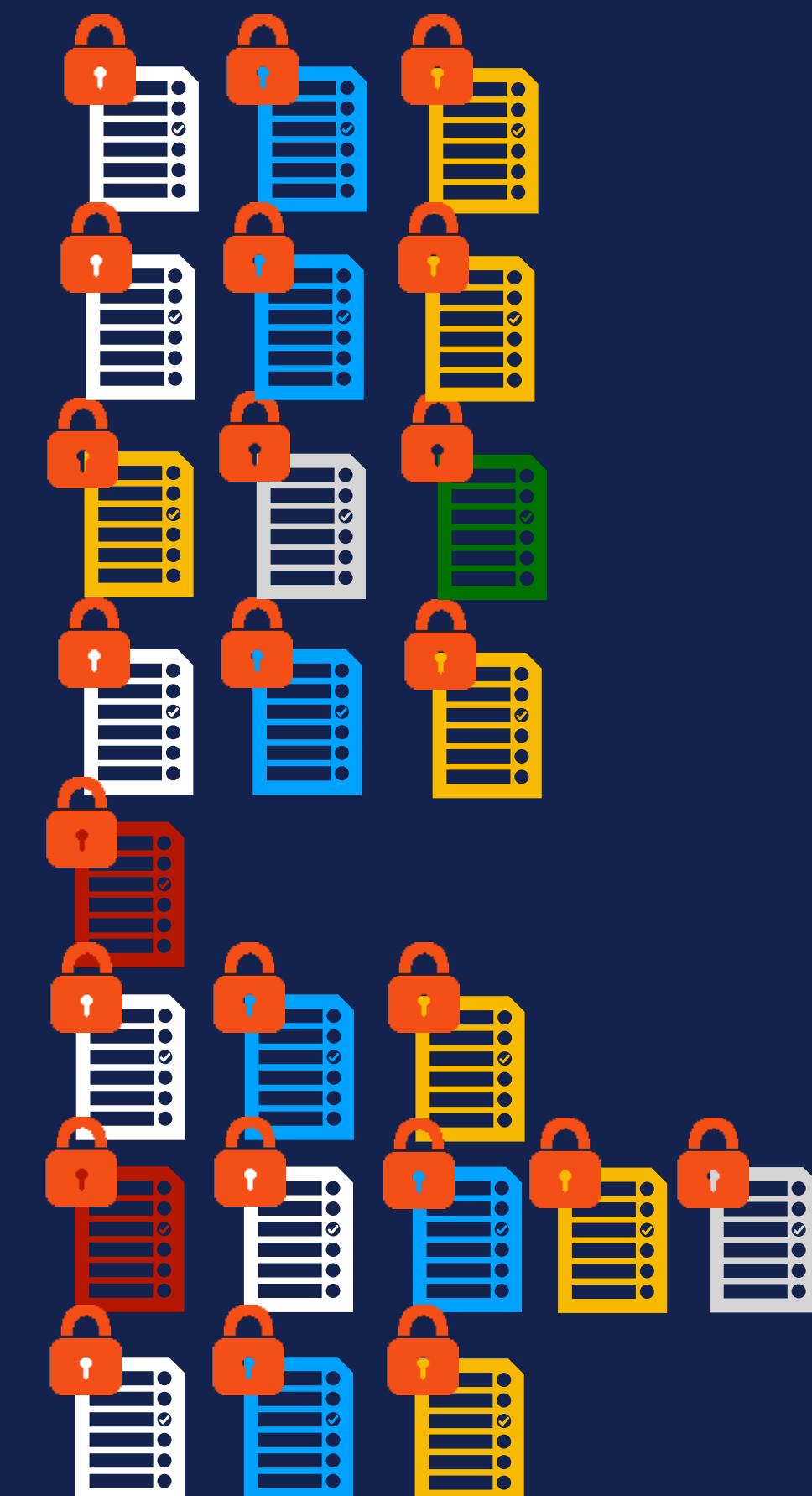
$\text{PRF}_K((\bullet, \circ))$

$\text{PRF}_K((\bullet, \bullet))$

$\text{PRF}_K((\circ, \bullet))$

$\text{PRF}_K((\bullet, \circ))$

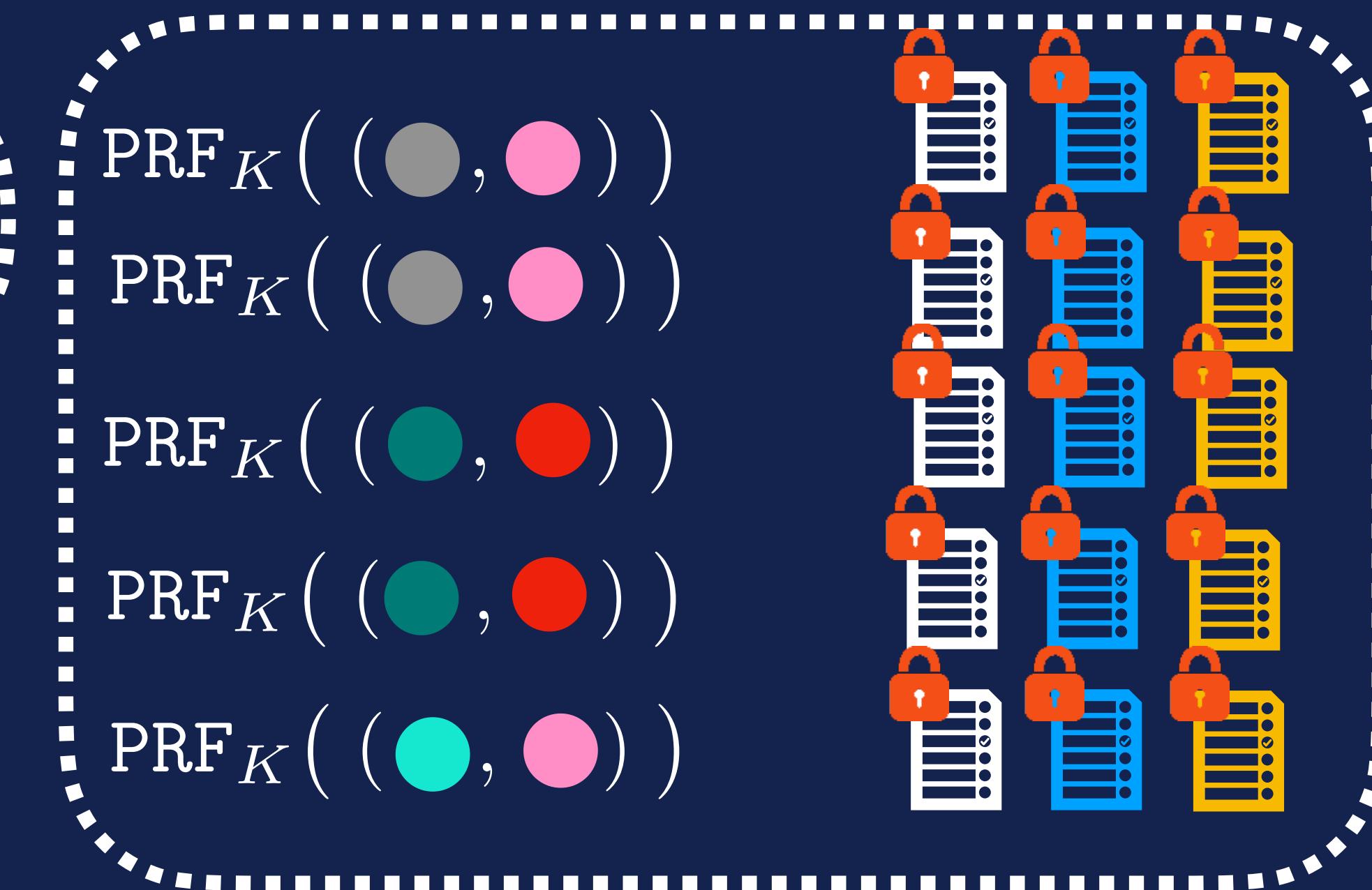
$\text{PRF}_K((\bullet, \bullet))$





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

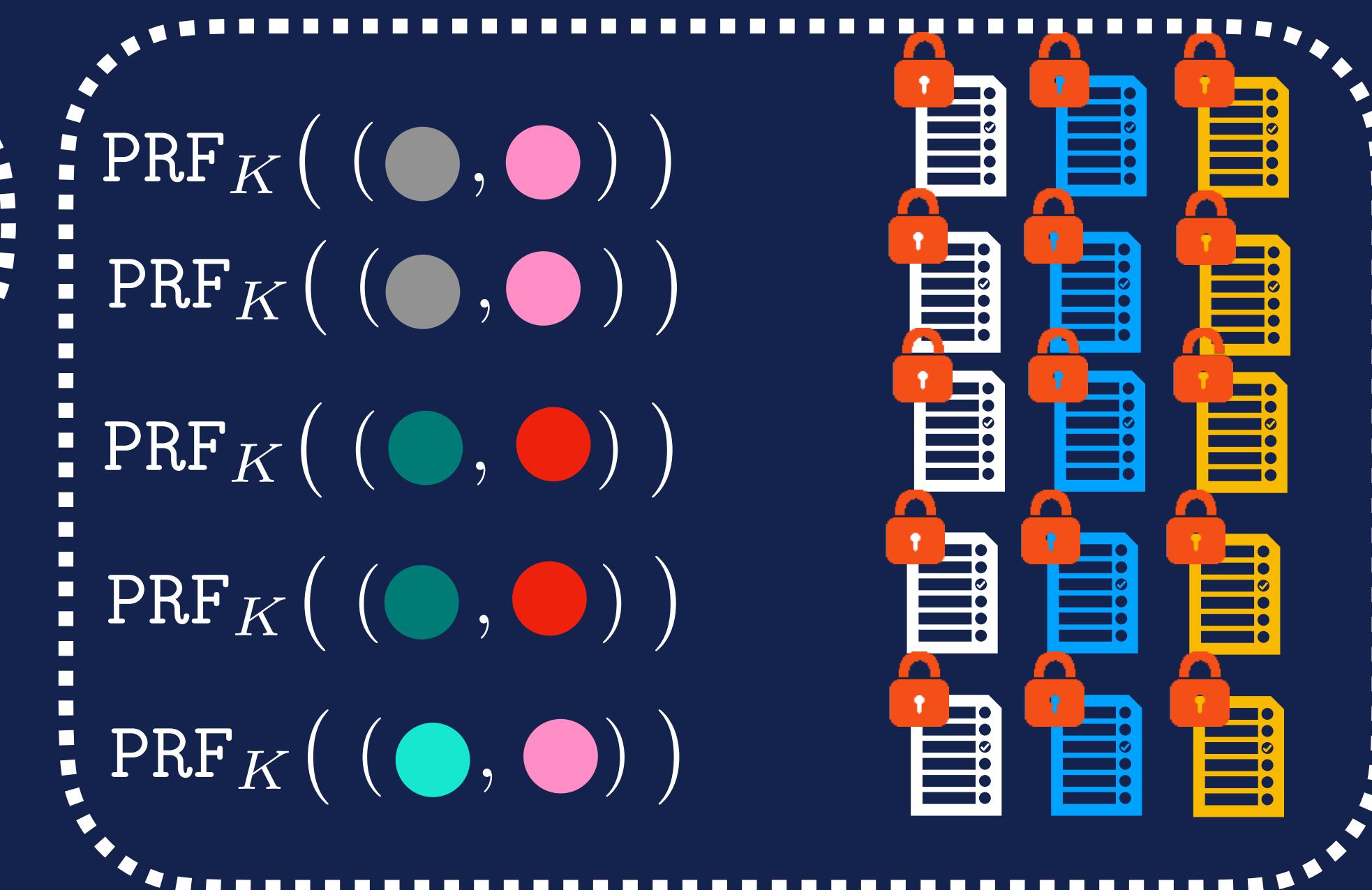
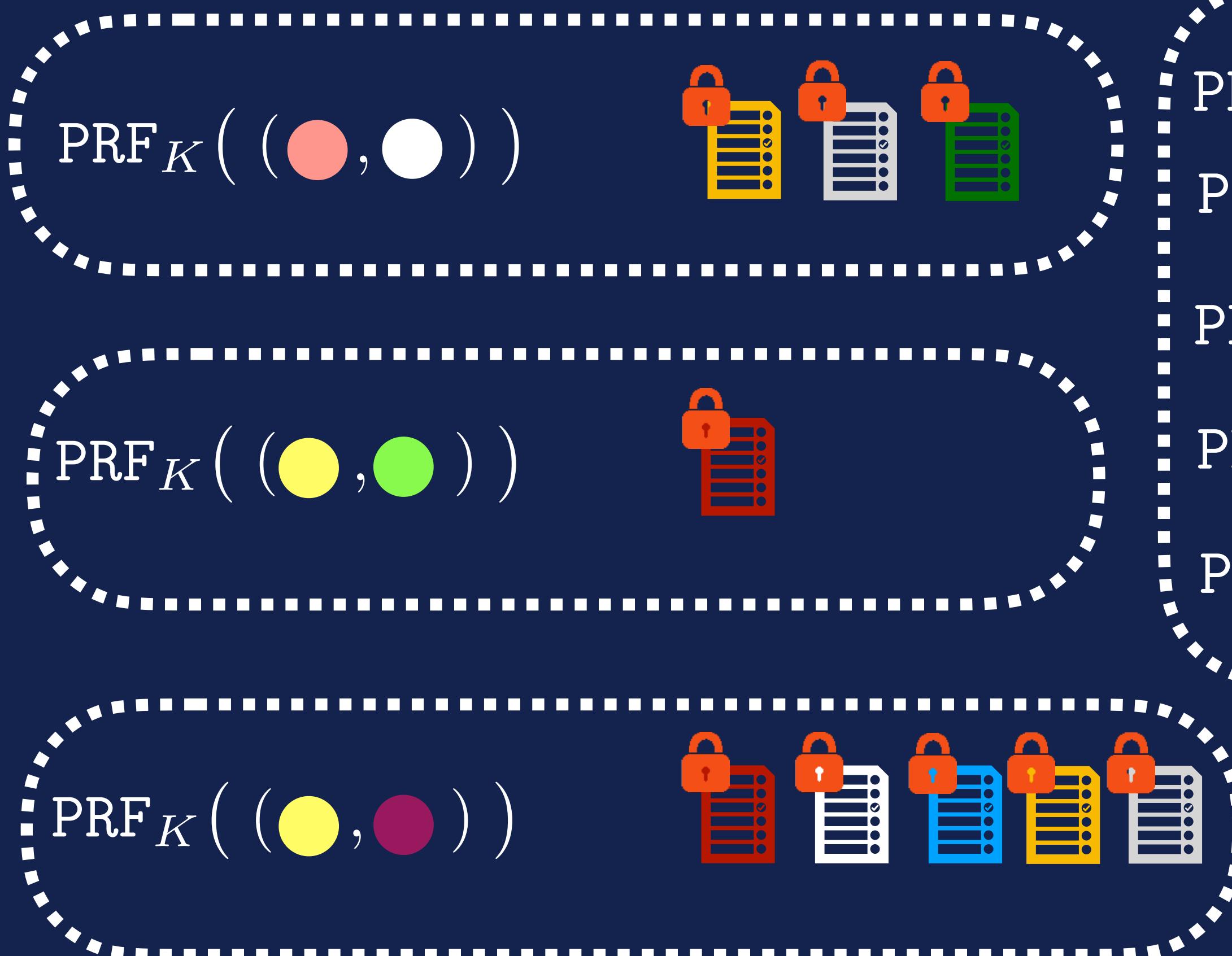
- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

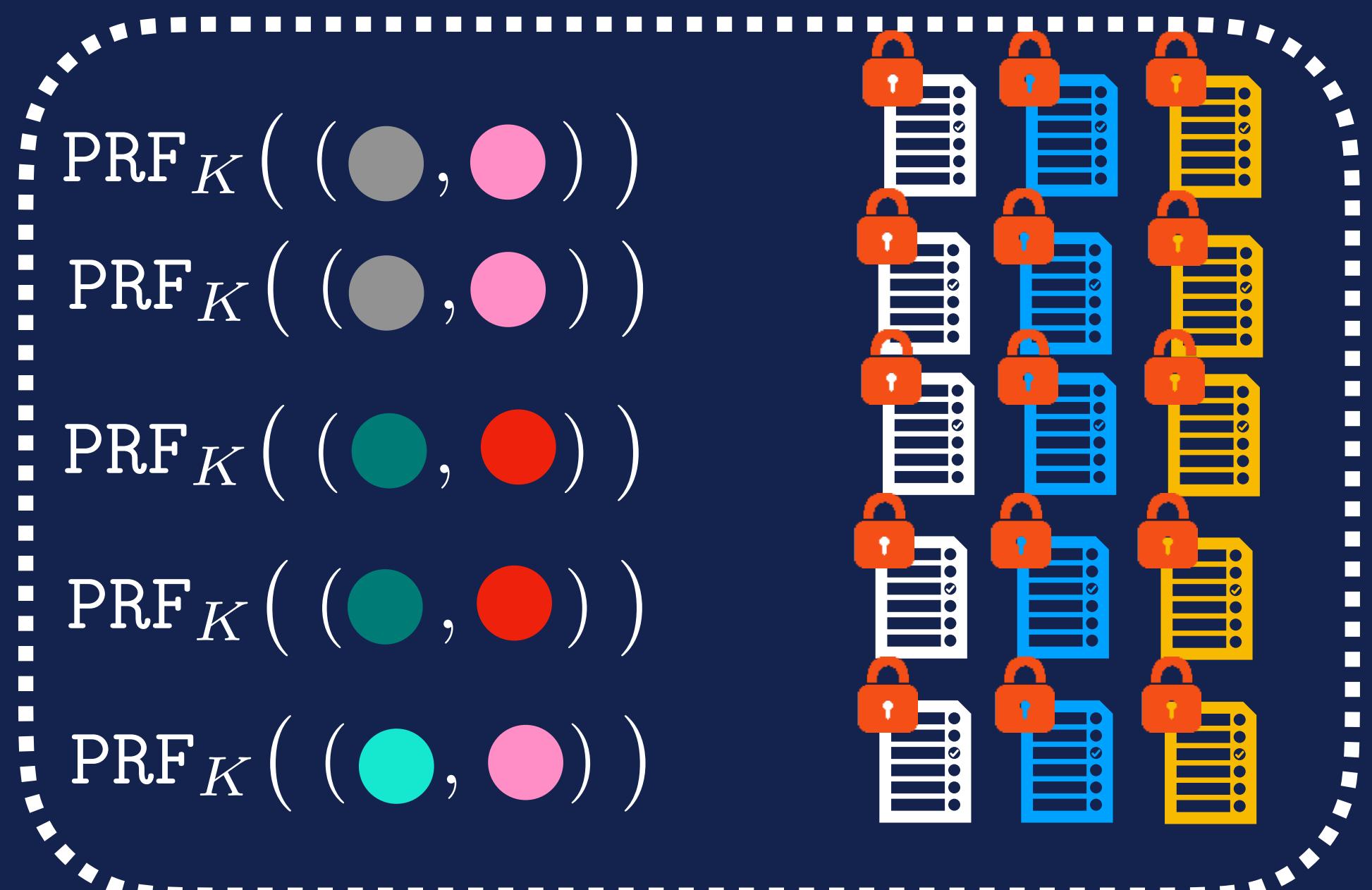
- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

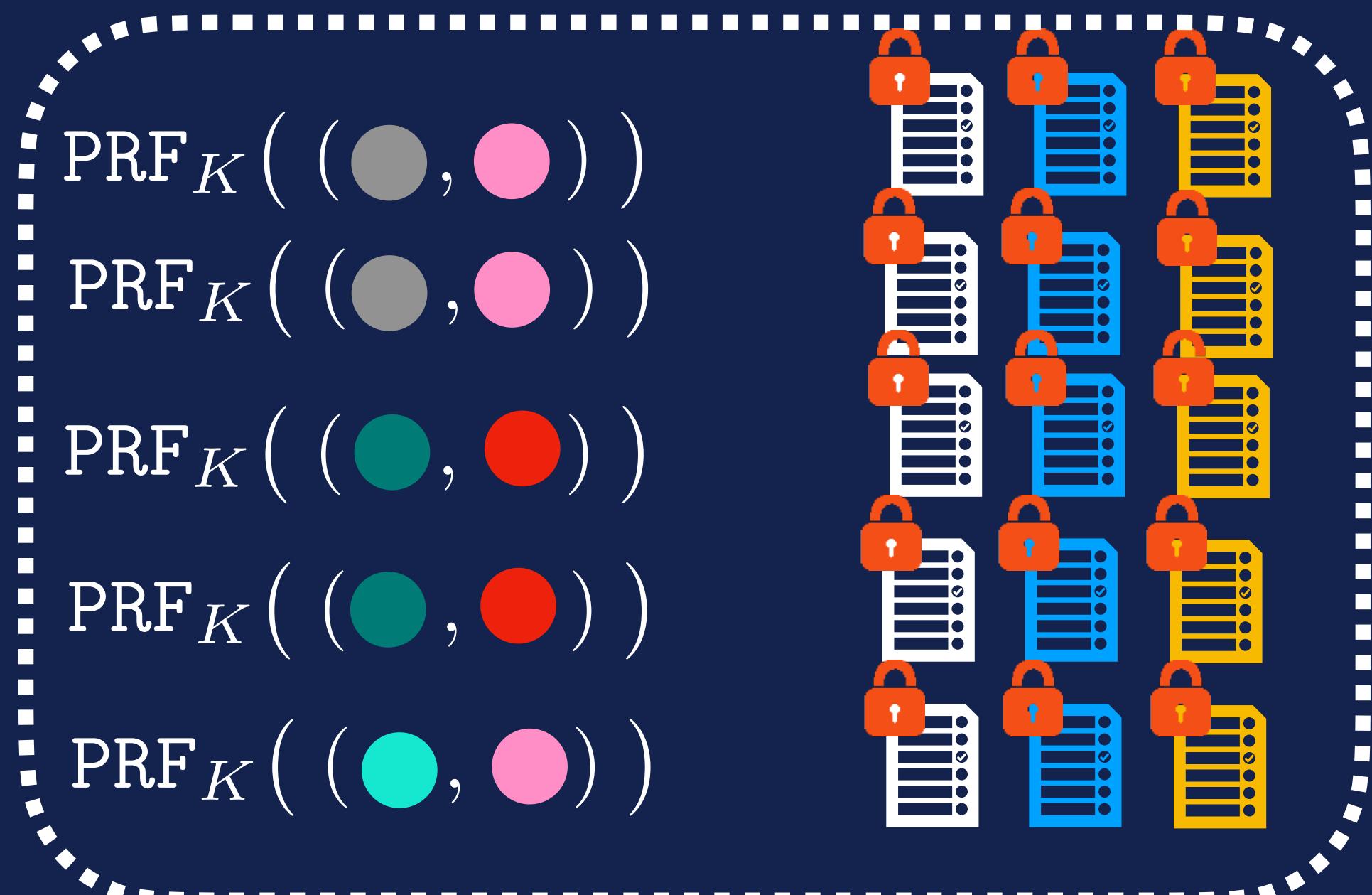
- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response

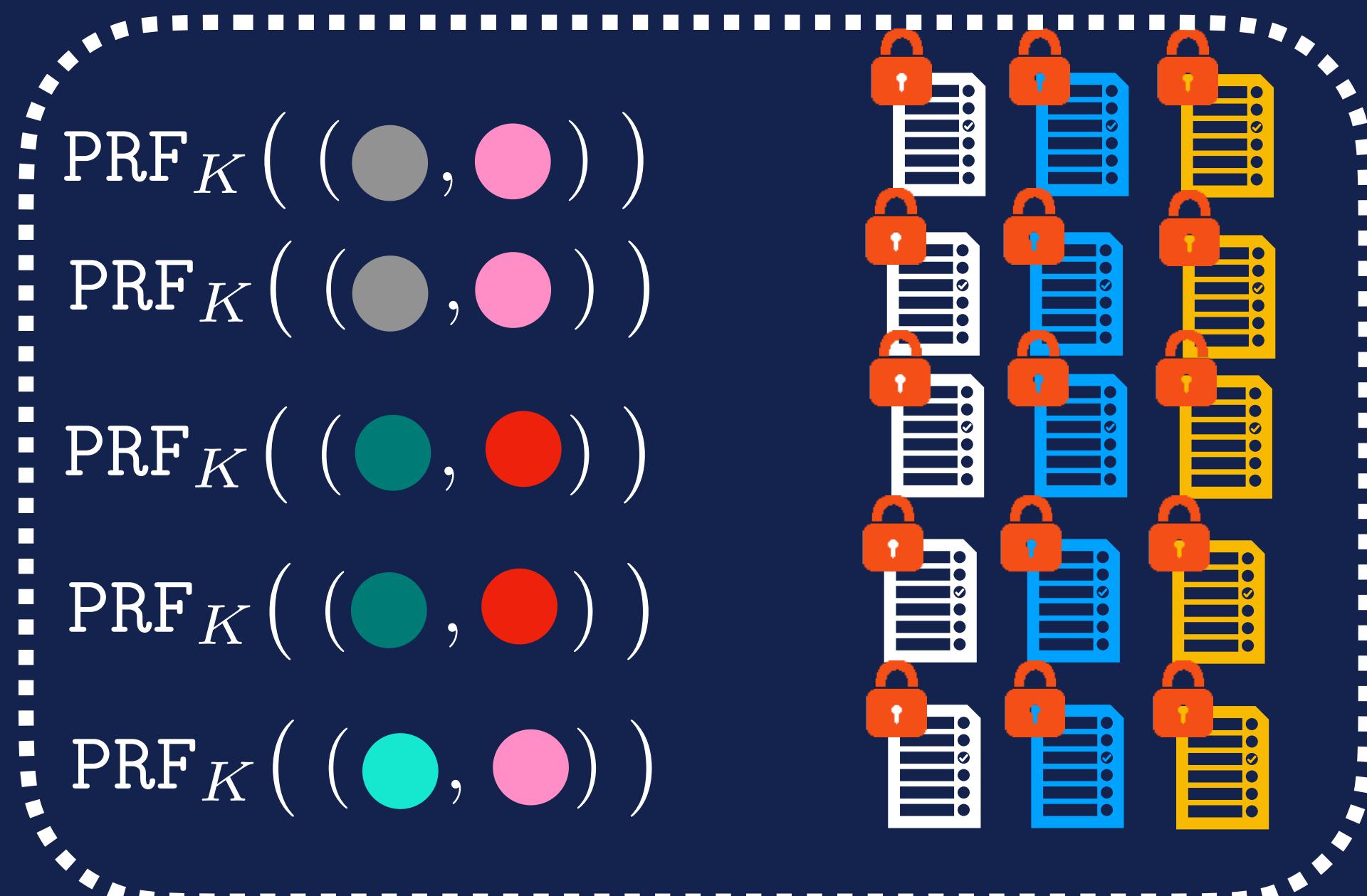


How many **distinct tokens** exist
that return response  ?



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response

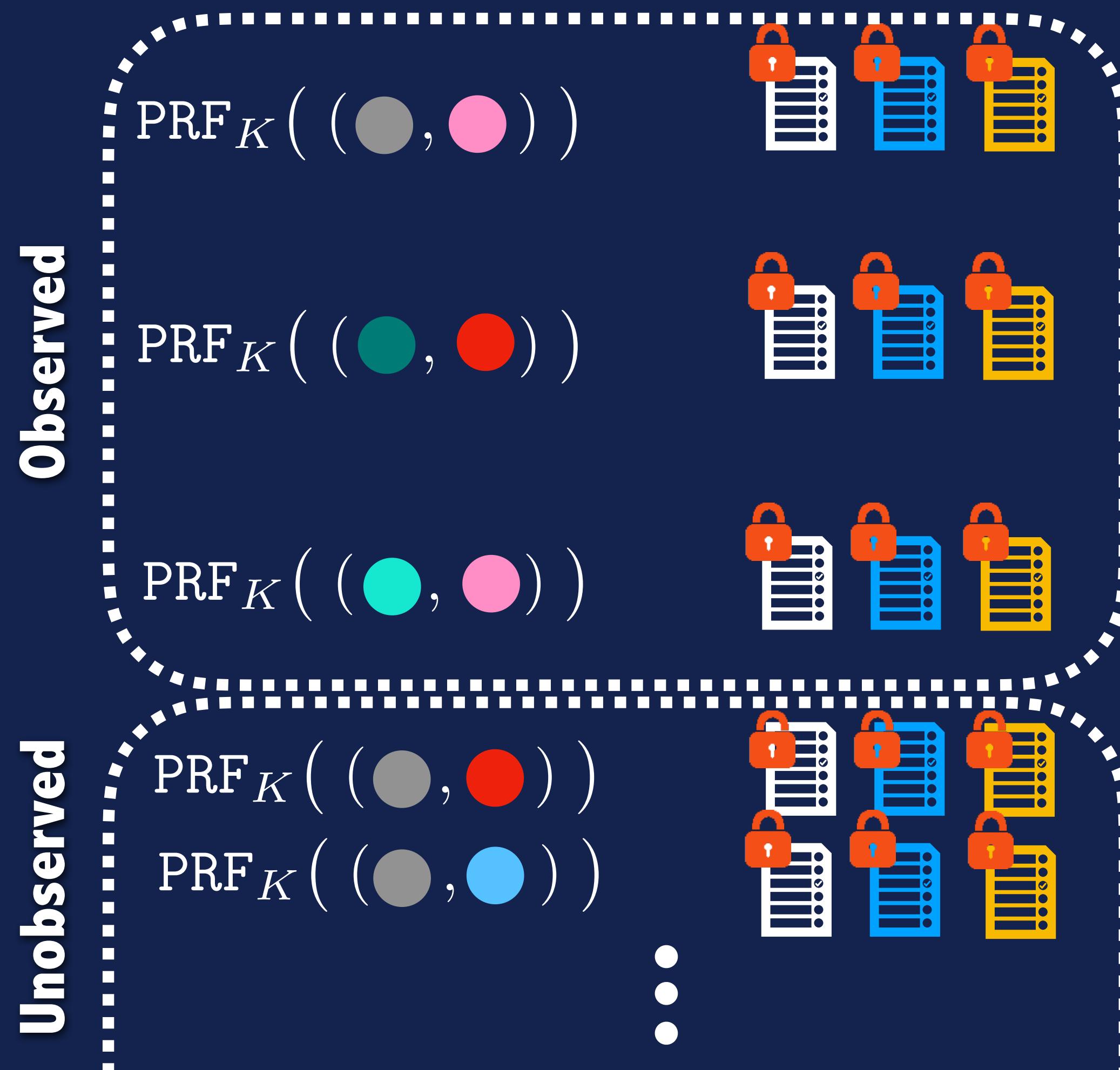


How many **distinct tokens** exist
that return response  ?



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response



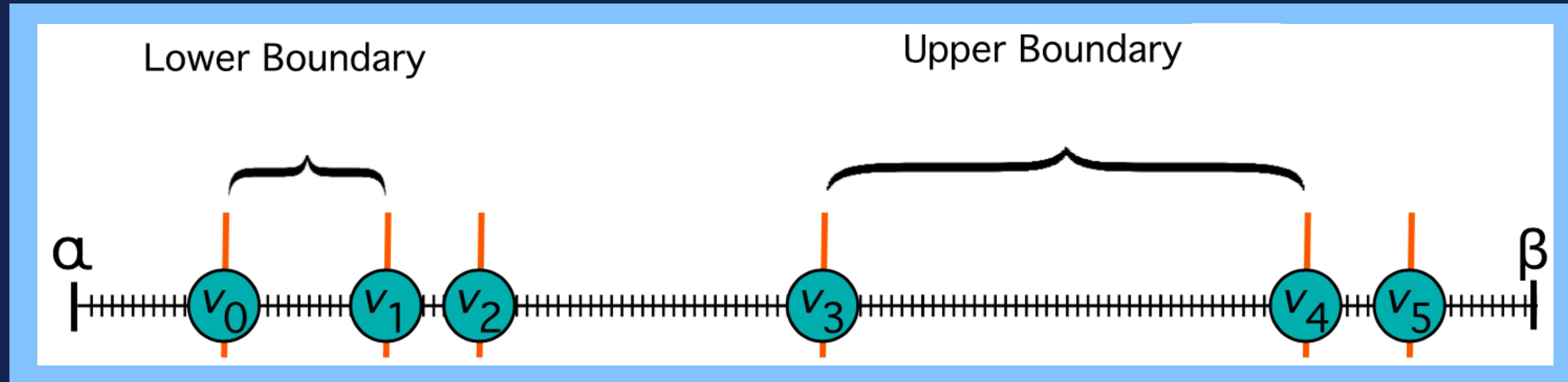
How many **distinct tokens** exist
that return response ?





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

Plaintext “Universe”

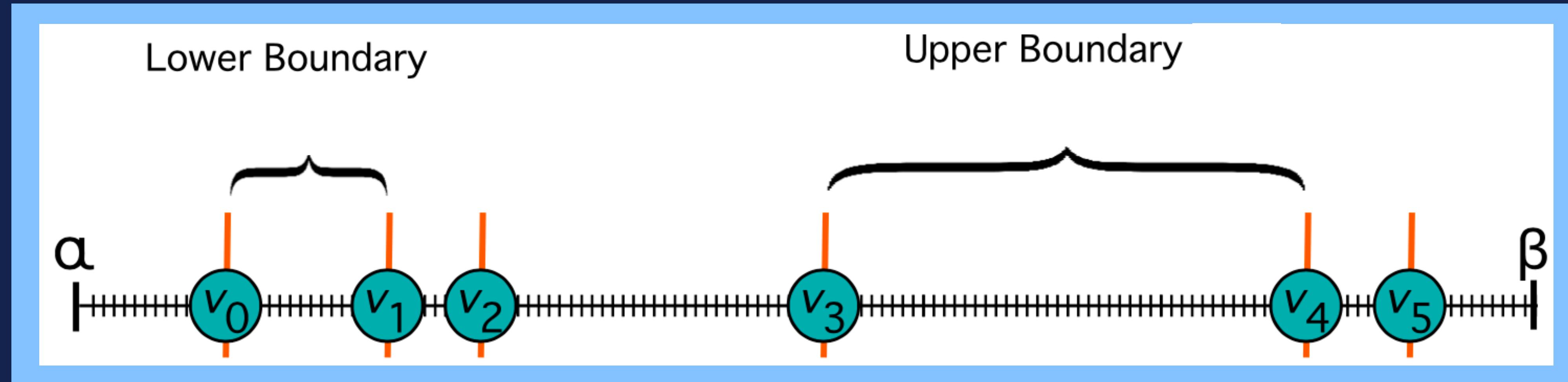


Number of **distinct range queries with response** v_1, v_2, v_3 ?



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

Plaintext “Universe”



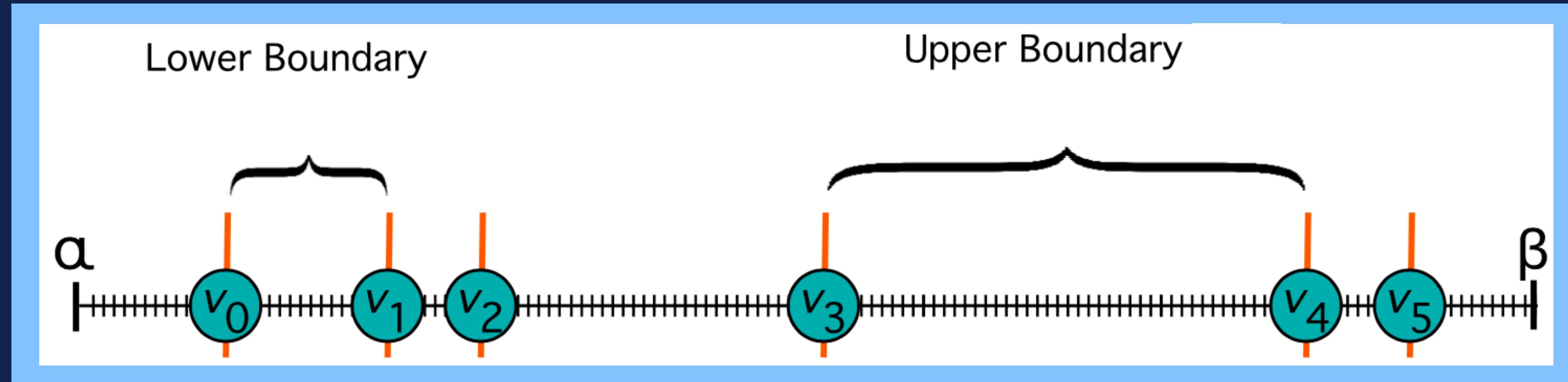
Number of **distinct range queries with response**  ?

$$|Q_r| = d(v_0, v_1) \cdot d(v_3, v_4)$$



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

Plaintext “Universe”



Number of **distinct range queries with response**  ?

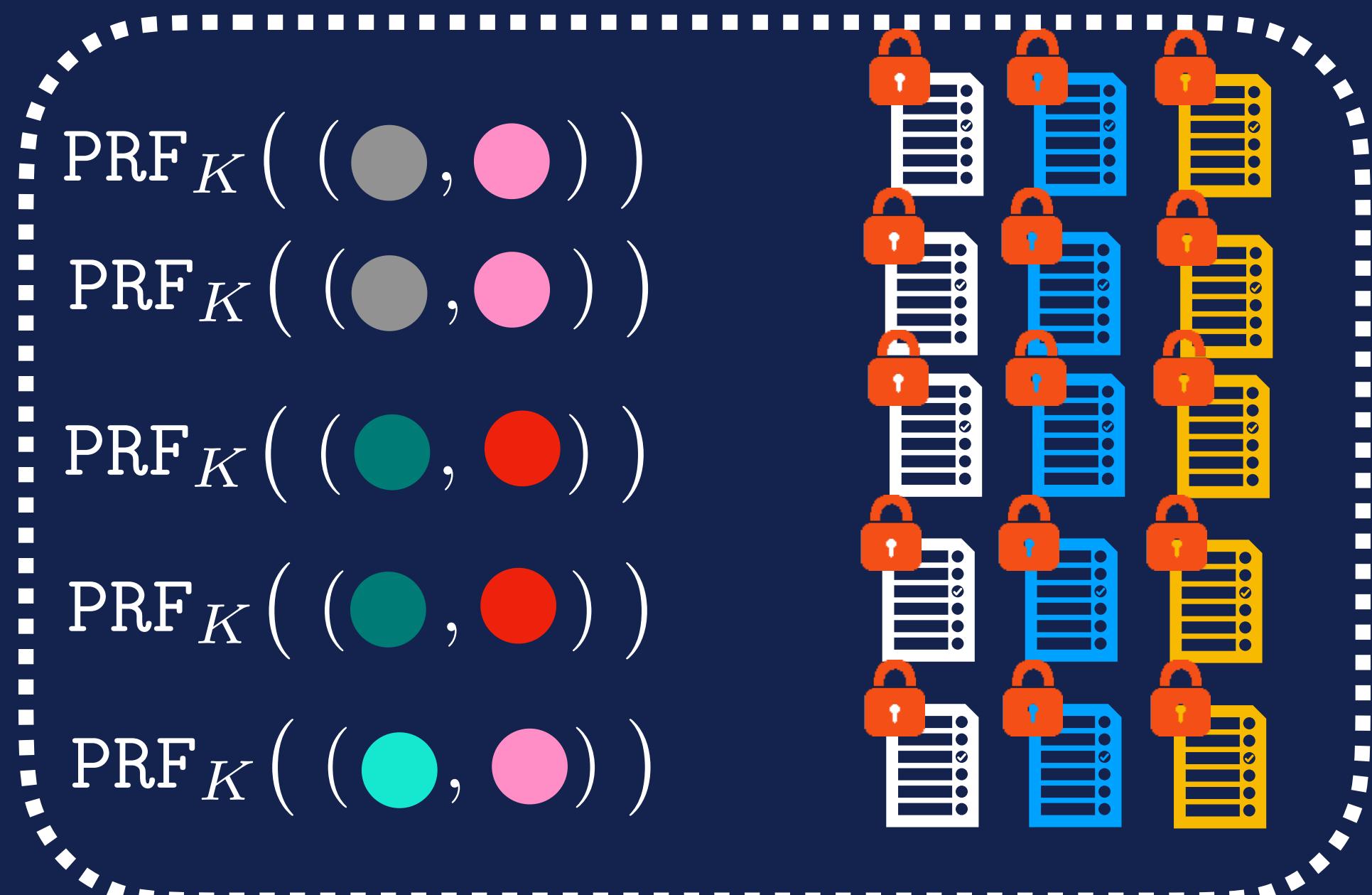
$$|Q_r| = d(v_0, v_1) \cdot d(v_3, v_4)$$

If we infer the number of **distinct range queries** we learn the **product of distances!**



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response

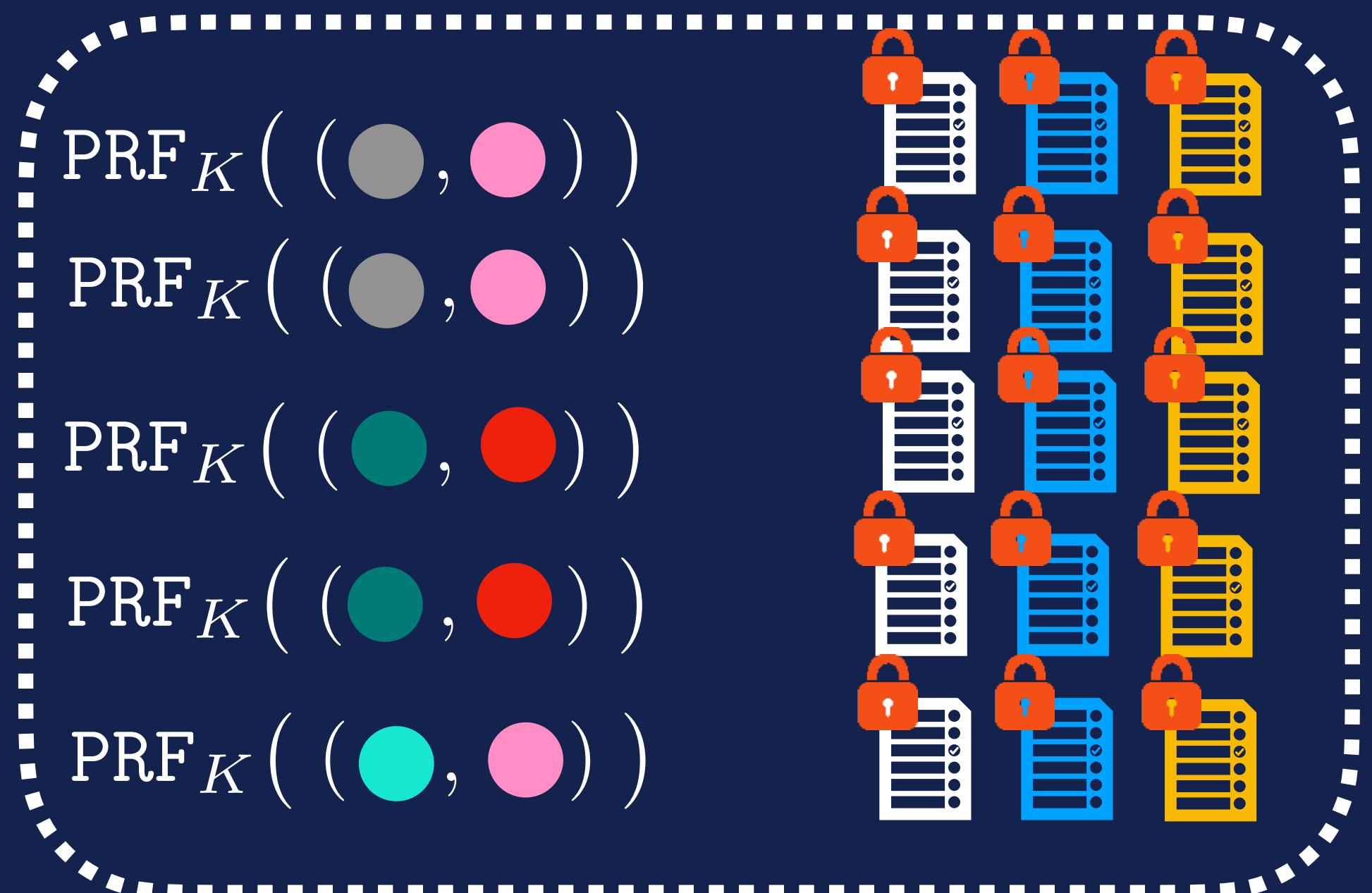


How many **distinct tokens** exist
that return response  ?



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

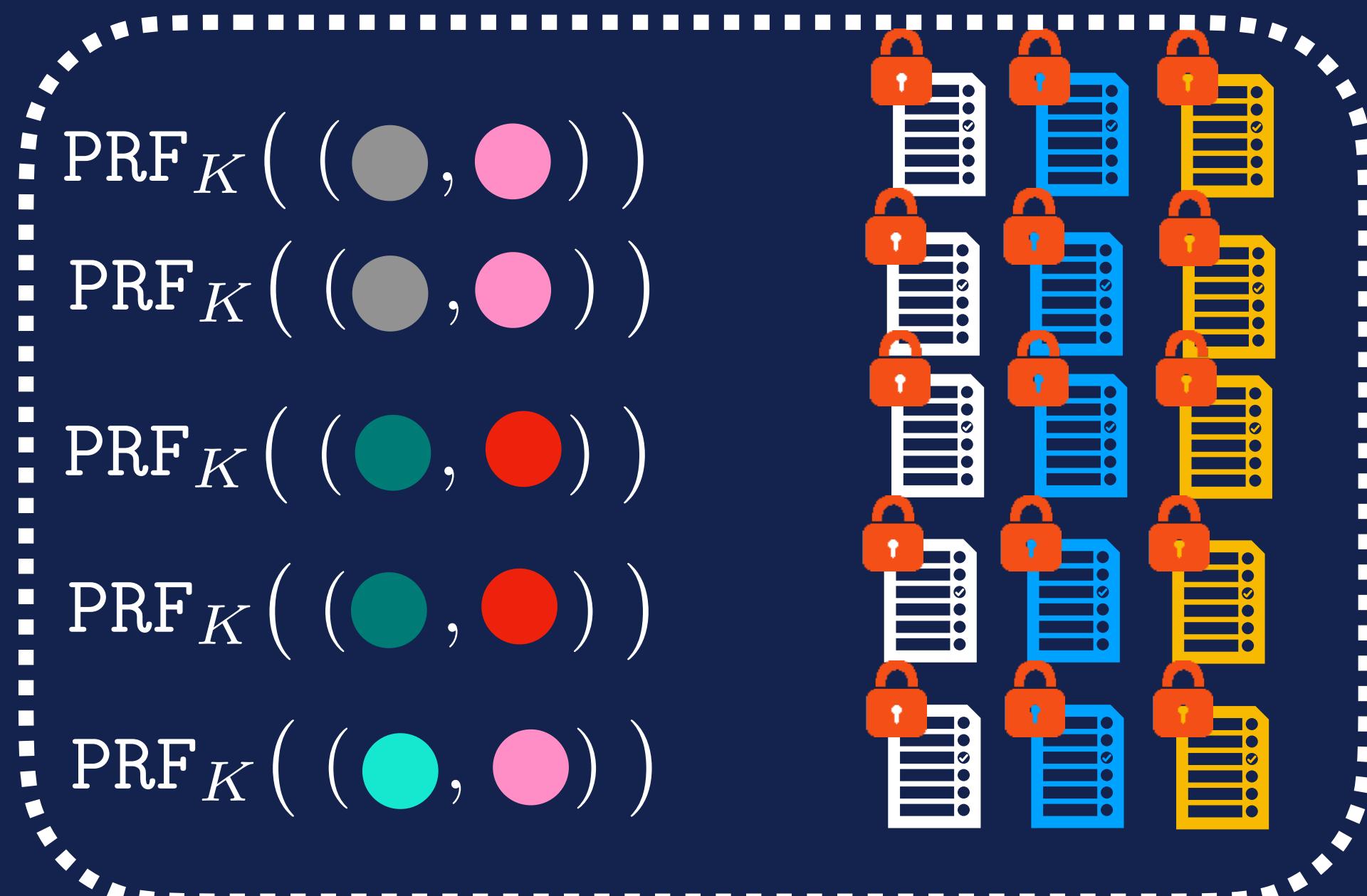
- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response



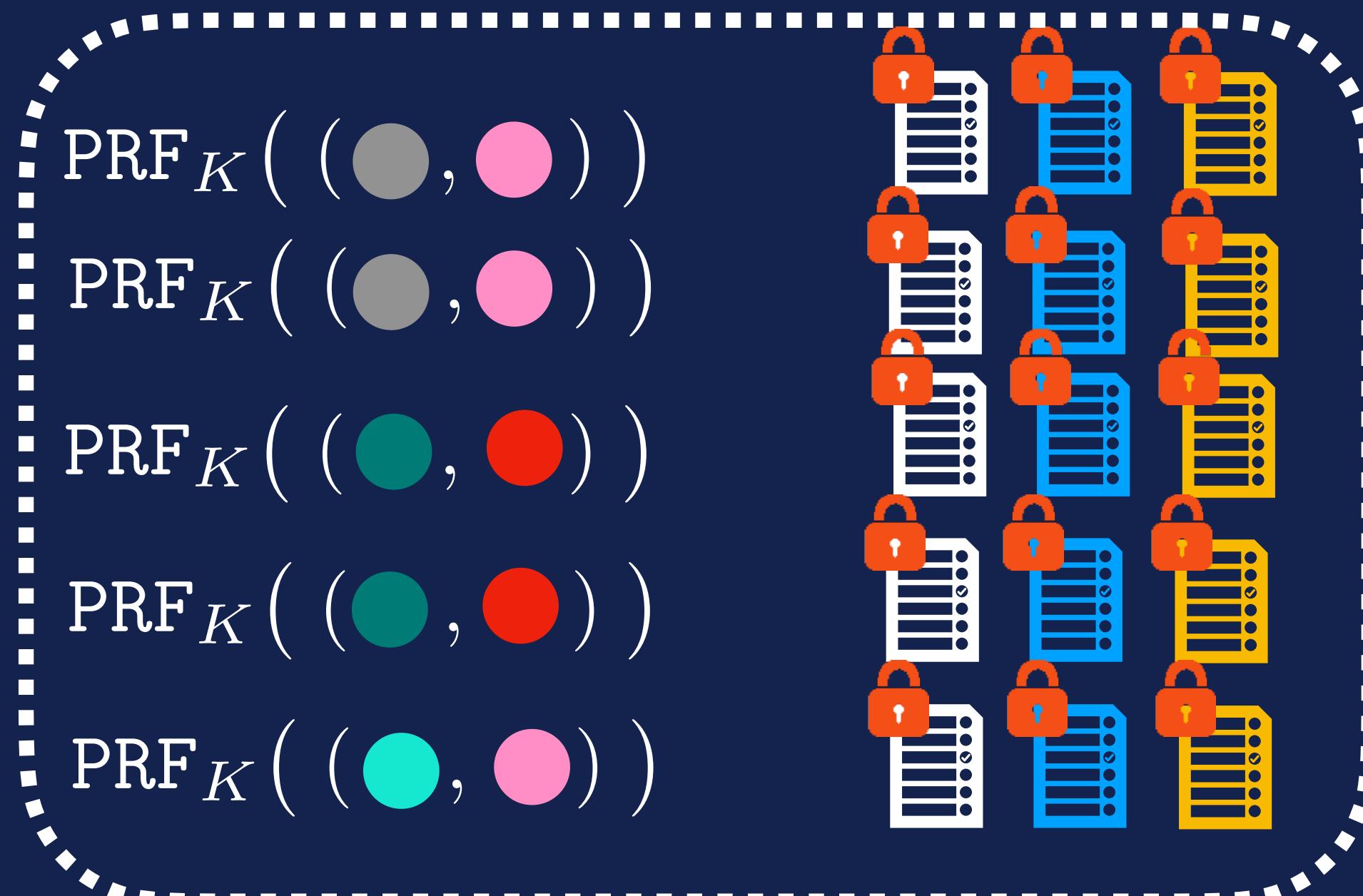
Can we estimate how many distinct tokens exist that return response ?





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response



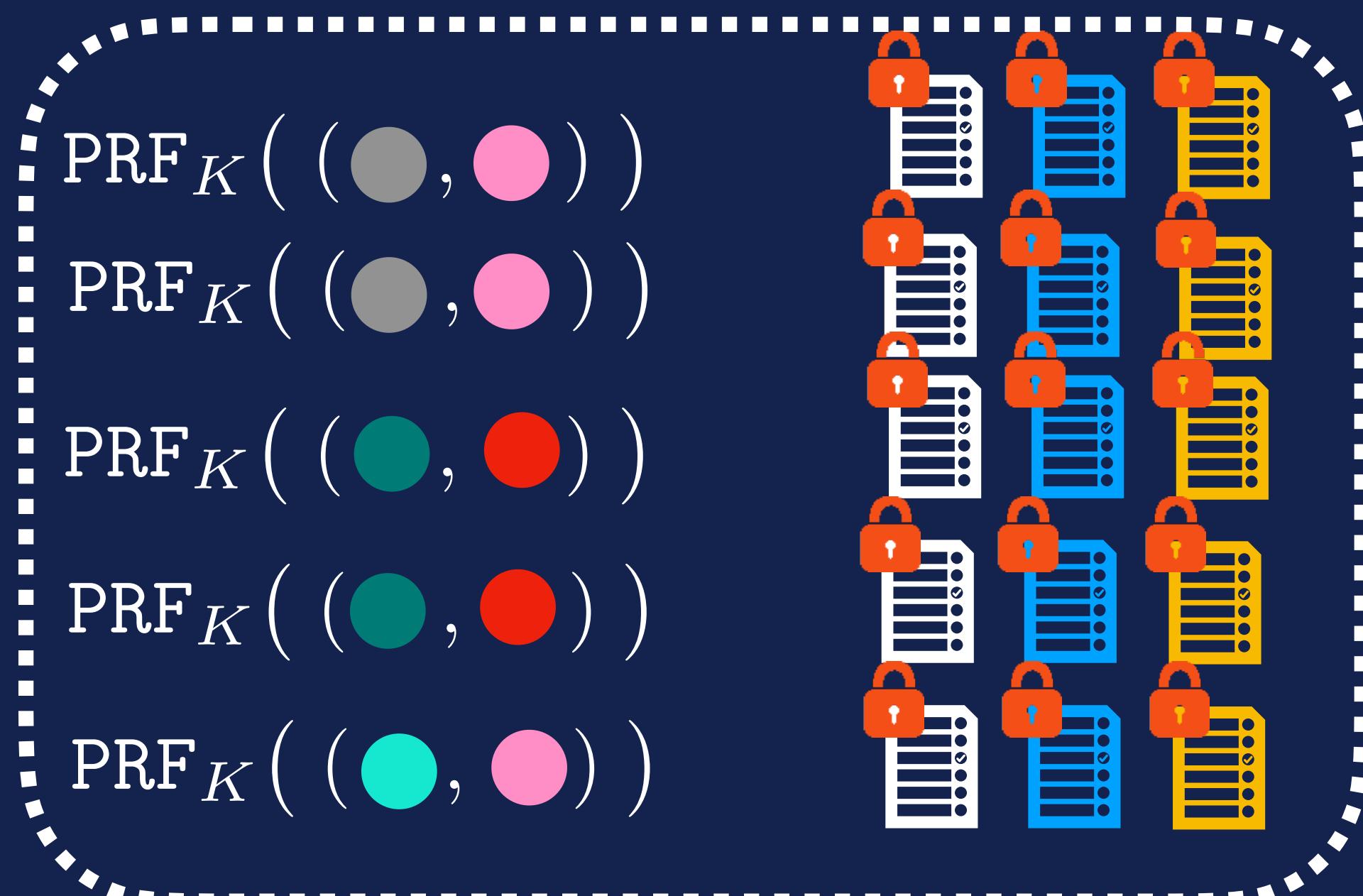
Can we estimate how many distinct tokens exist that return response ?

T: Random variable takes values from universe of tokens
R: Random variable takes values from universe of Responses



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response



Sample from Pr(T | R= { , , })



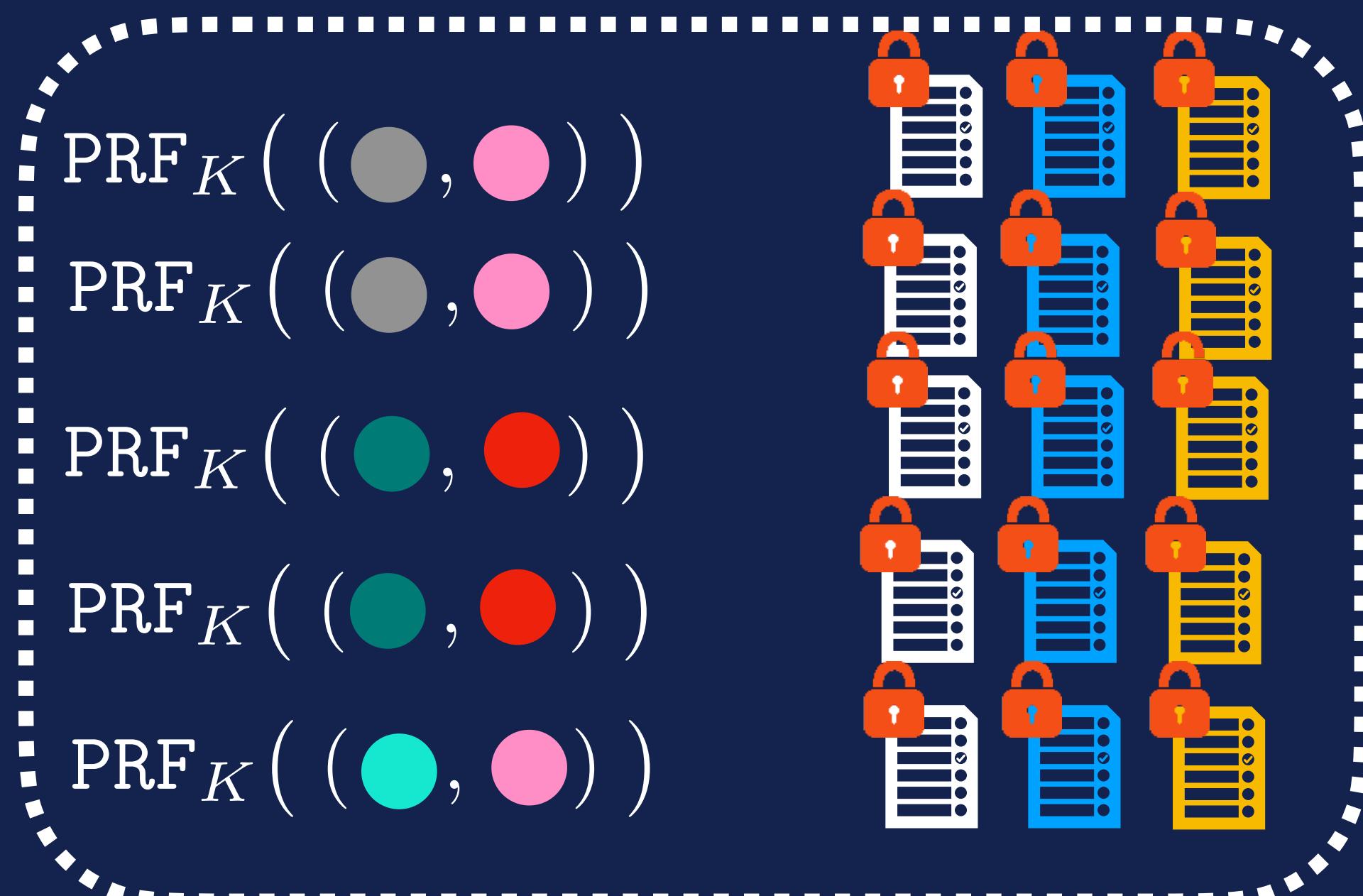
Can we estimate how many distinct tokens exist that return response ?

T: Random variable takes values from universe of tokens
R: Random variable takes values from universe of Responses



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response



Sample from $\Pr(T | R = \{ \text{server icon} \})$



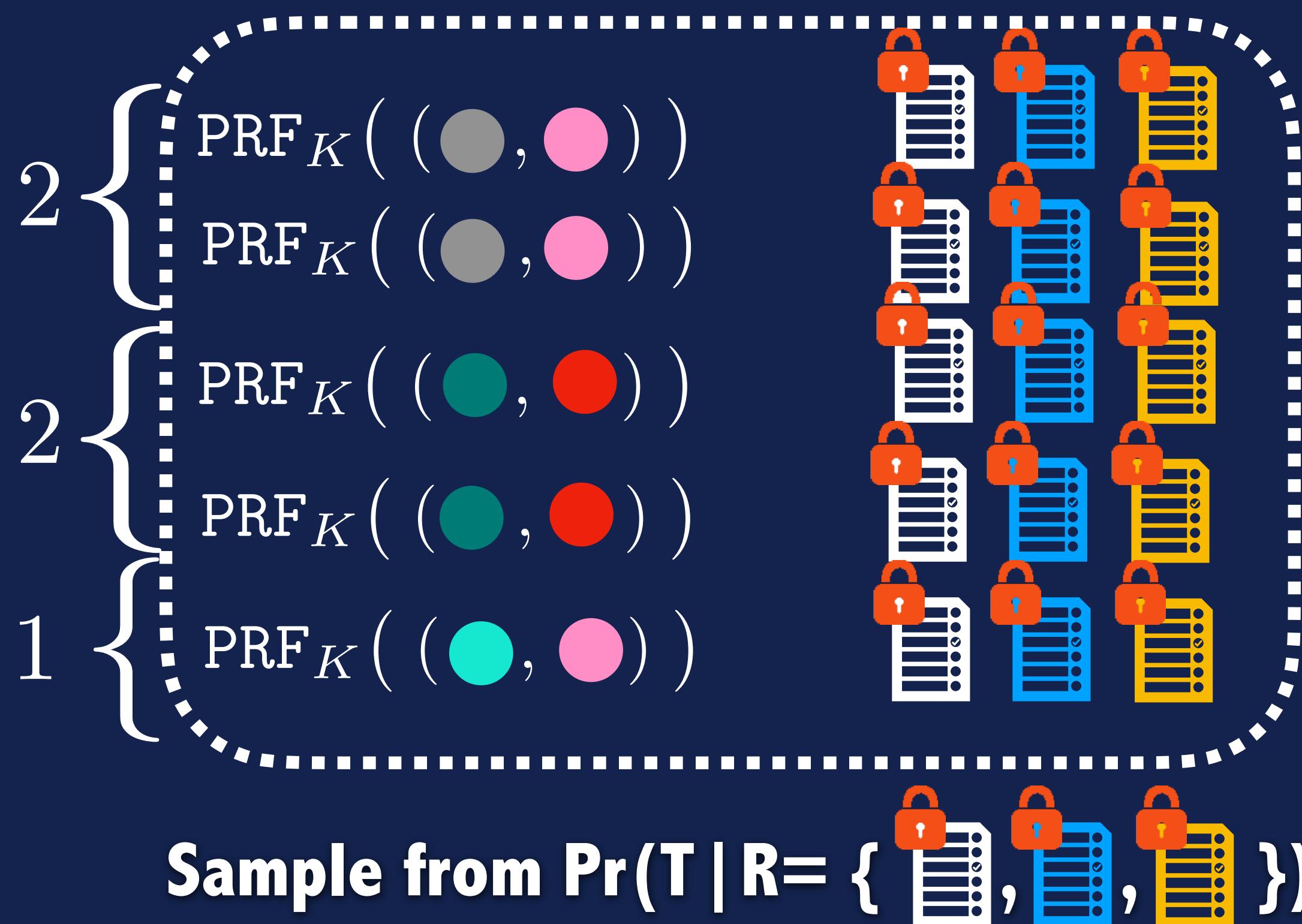
Can we estimate how many distinct tokens exist that return response ?

T: Random variable takes values from universe of tokens
R: Random variable takes values from universe of Responses



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response



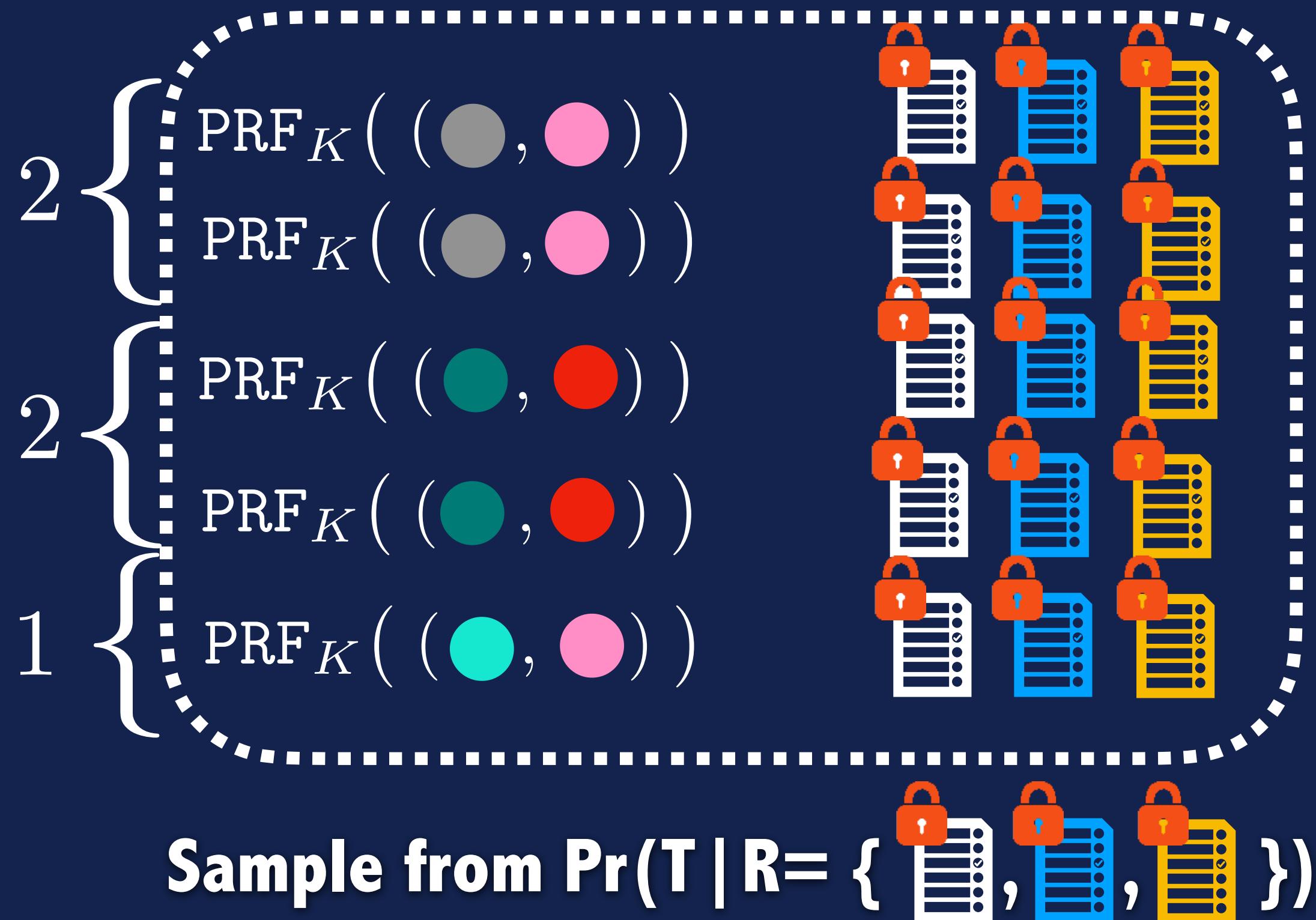
Can we estimate how many distinct tokens exist that return response ?

T: Random variable takes values from universe of tokens
R: Random variable takes values from universe of Responses



STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

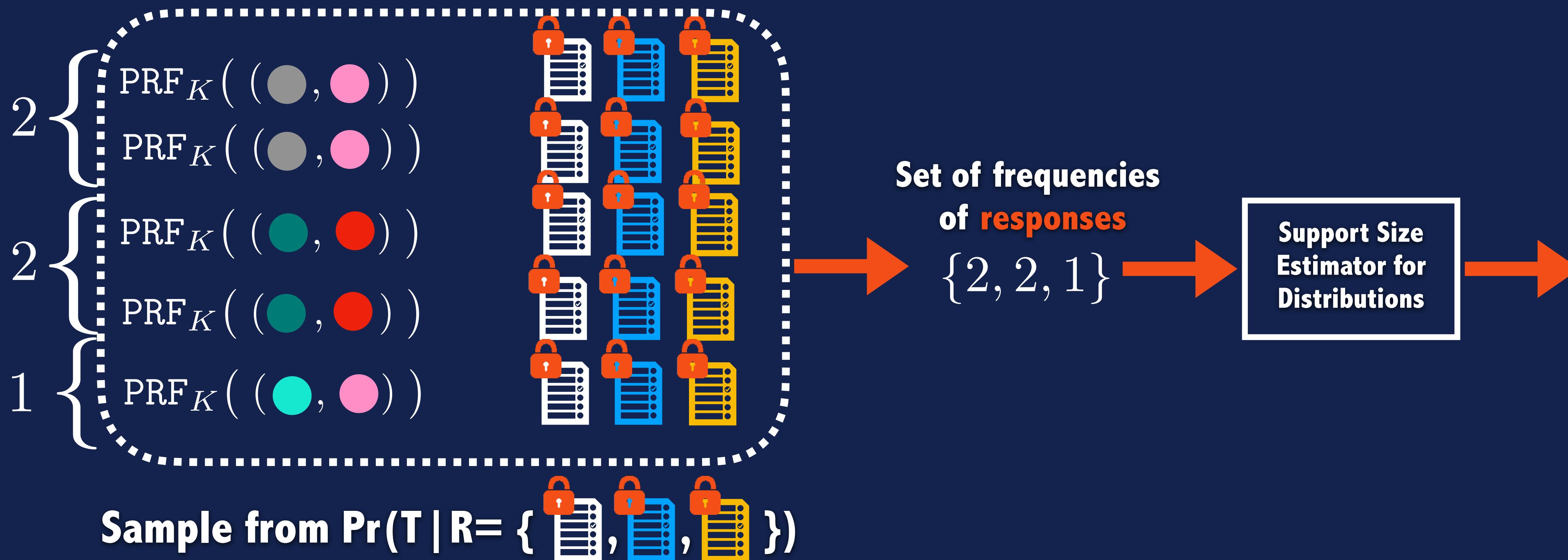
- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

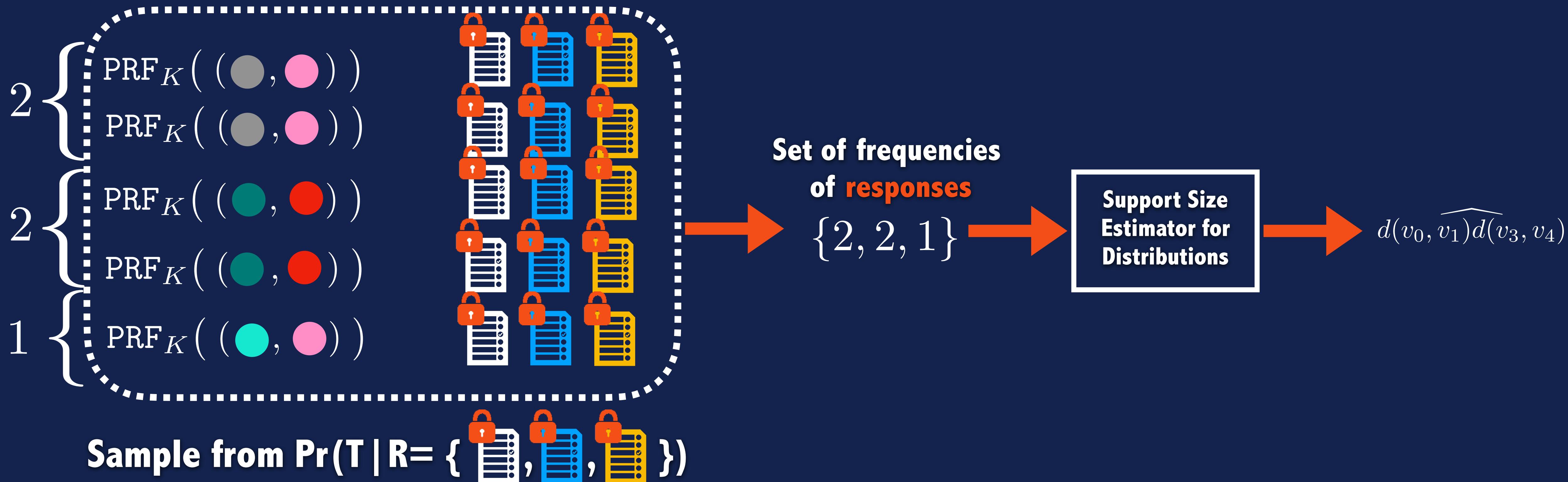
- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response





STATE OF THE UNIFORM SEARCH PATTERN + ACCESS PATTERN

- Consider a token as an “encrypted pair of boundaries”
- Partition the token-response with respect to the response





LET'S TALK ABOUT SUPPORT SIZE ESTIMATORS

JACKKNIFE

- Non-parametric
- Frequency of each token as input
- Based on bias reduction, order decided based on the sample

VALIANT-VALIANT

- Non-parametric
- Frequency of each token as input
- “Simplest Histogram”



WHAT CAN THE ADVERSARY LEARN FROM THE SEARCH PATTERN LEAKAGE ?



WHAT CAN THE ADVERSARY LEARN FROM THE SEARCH PATTERN LEAKAGE ?

Answer: From Frequencies of tokens we can estimate the product of pairwise distances

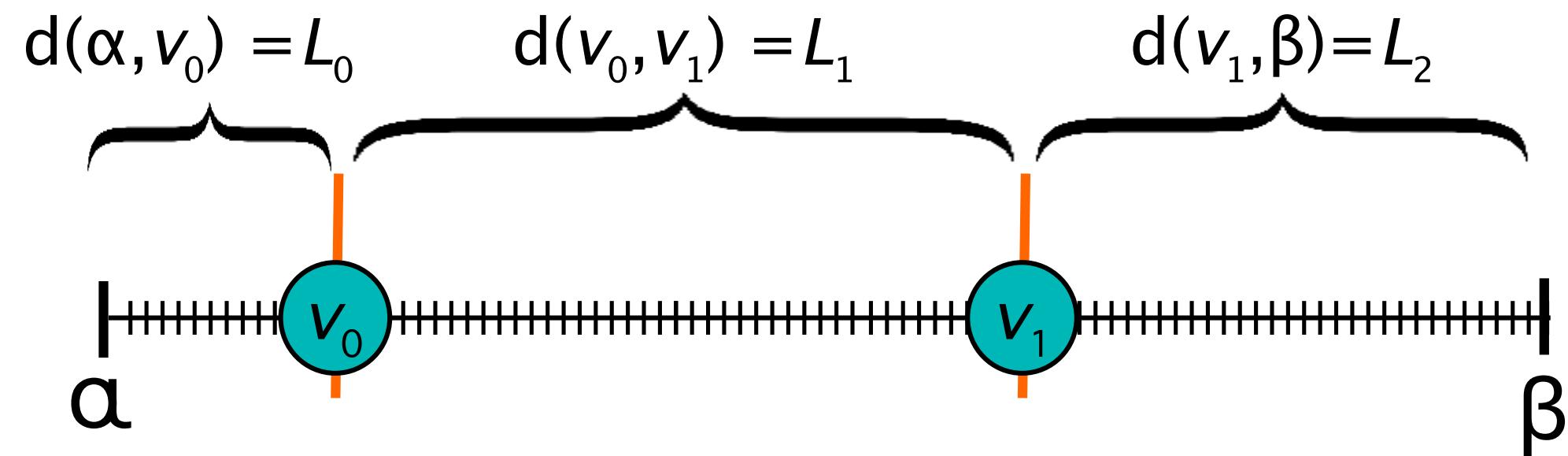


OVERVIEW OF THE ATTACK



OVERVIEW OF THE ATTACK

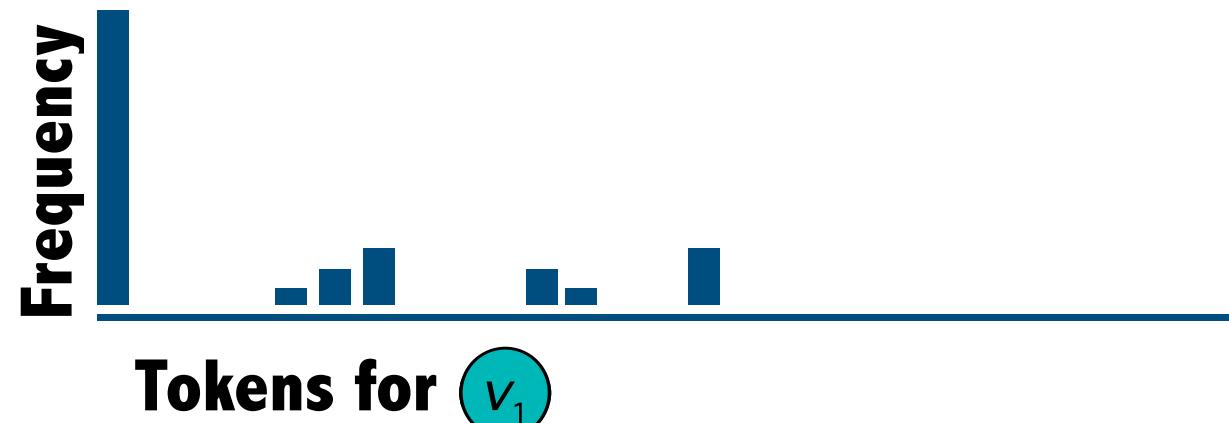
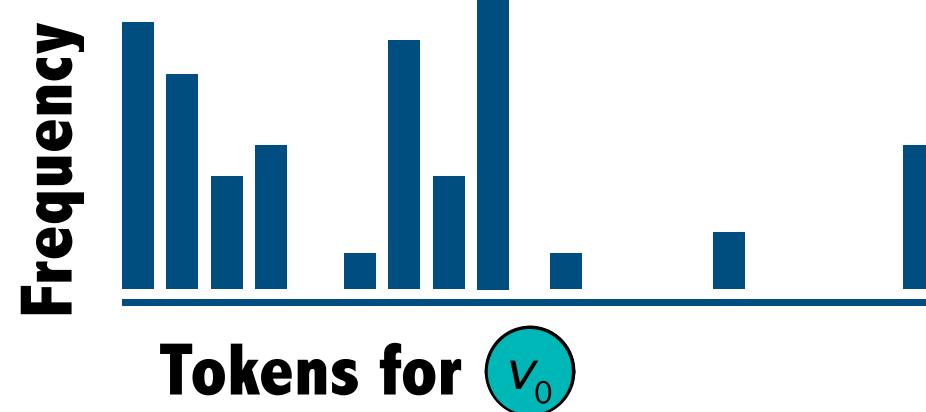
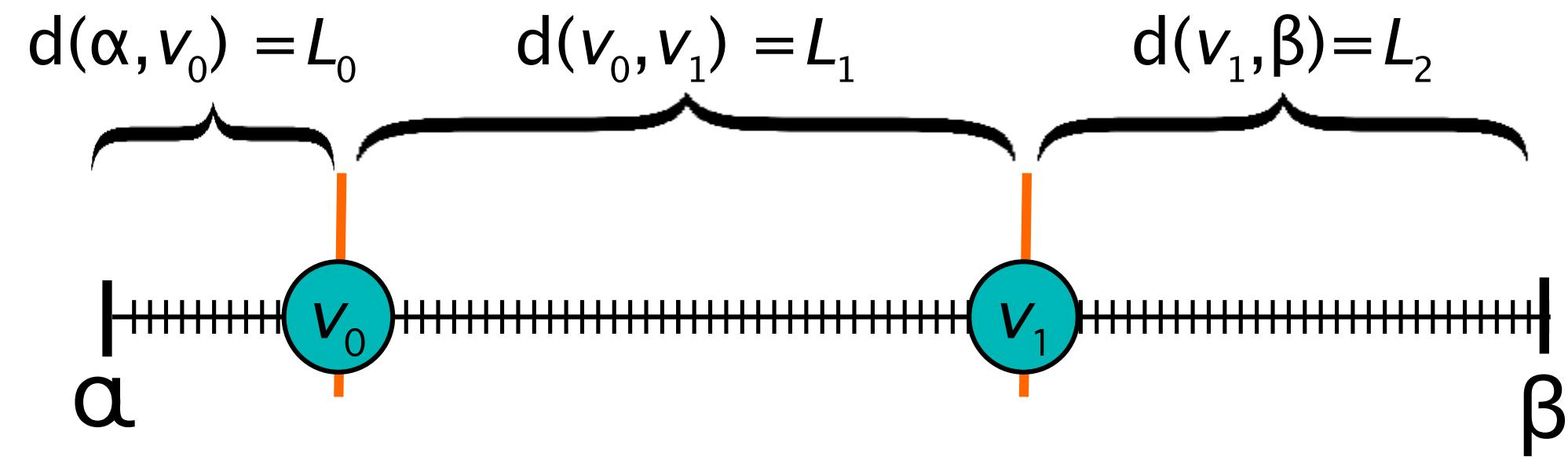
Plaintext:





OVERVIEW OF THE ATTACK

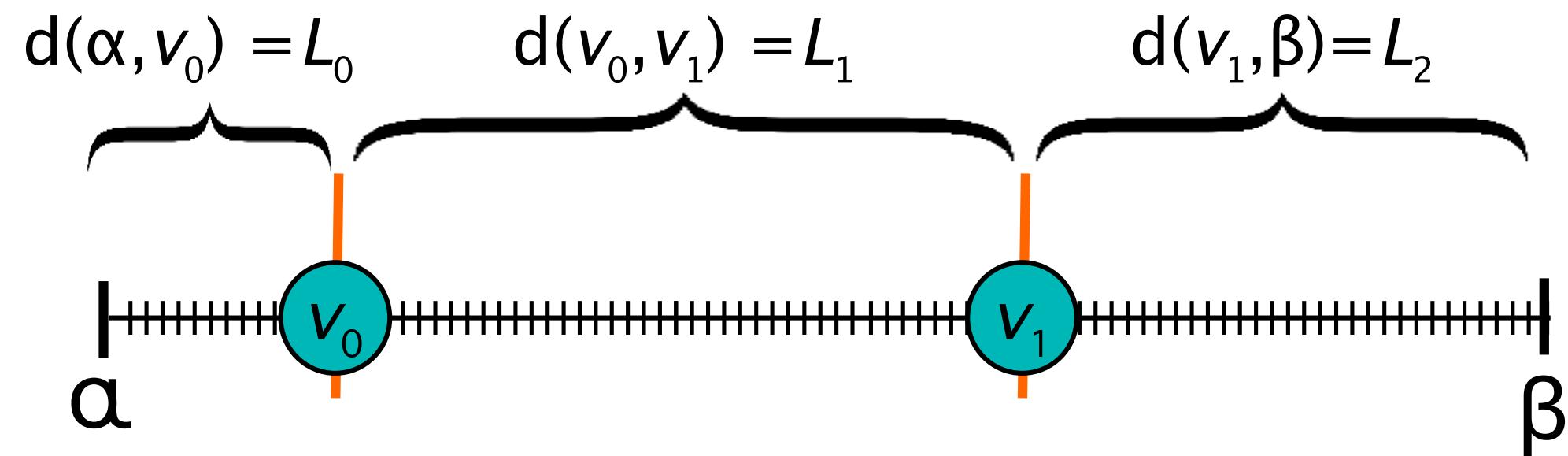
Plaintext:



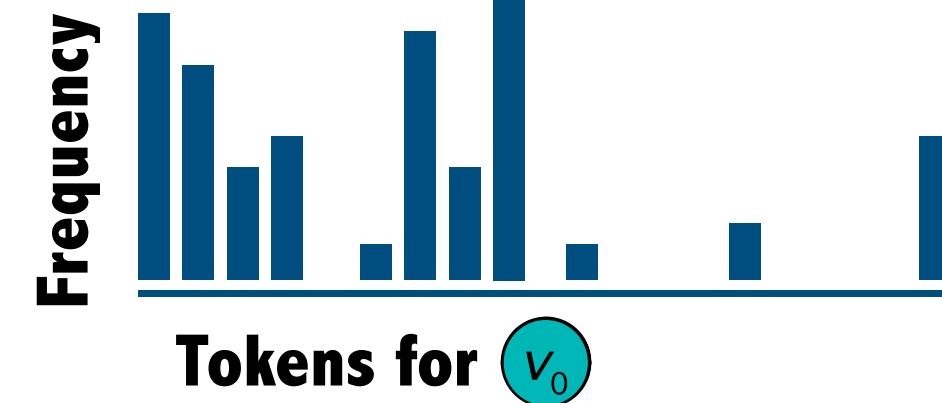


OVERVIEW OF THE ATTACK

Plaintext:



**Support Size Estimation
on Tokens**



$$\widehat{L}_0 \widehat{L}_1 = 350$$



$$\widehat{L}_1 \widehat{L}_2 = 1015$$

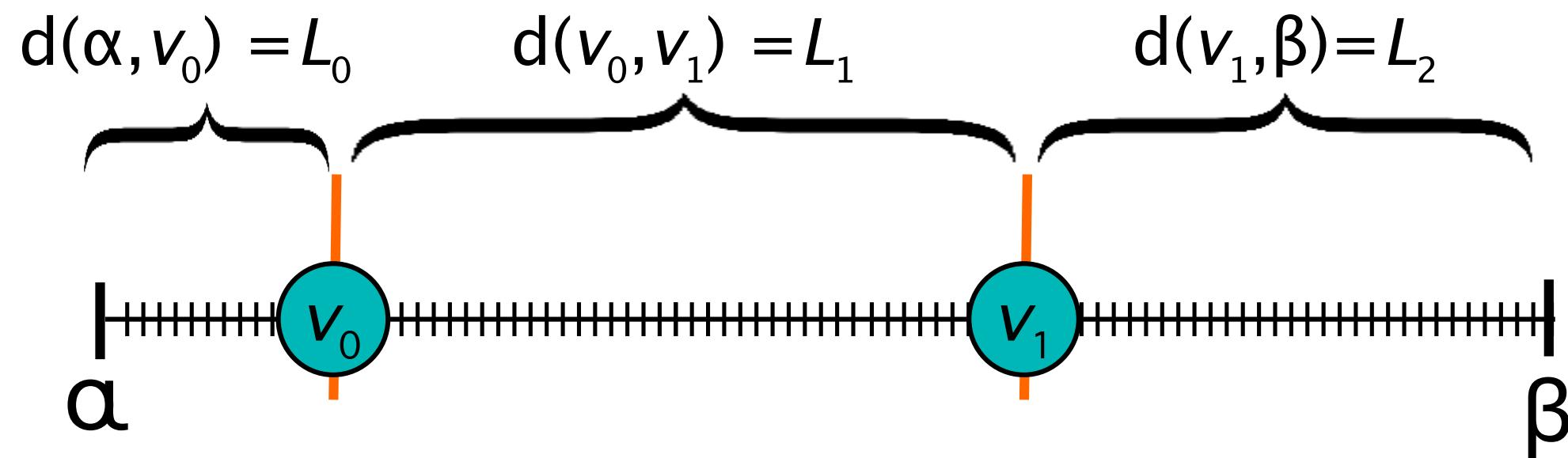


$$\widehat{L}_0 \widehat{L}_2 = 290$$

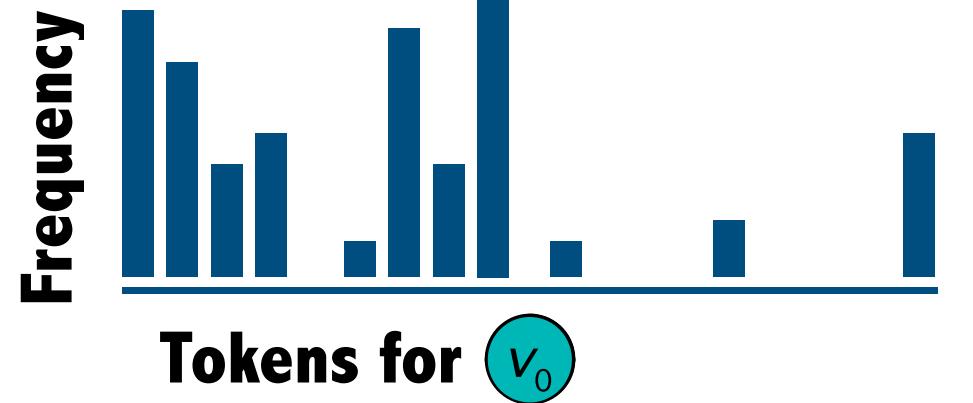


OVERVIEW OF THE ATTACK

Plaintext:

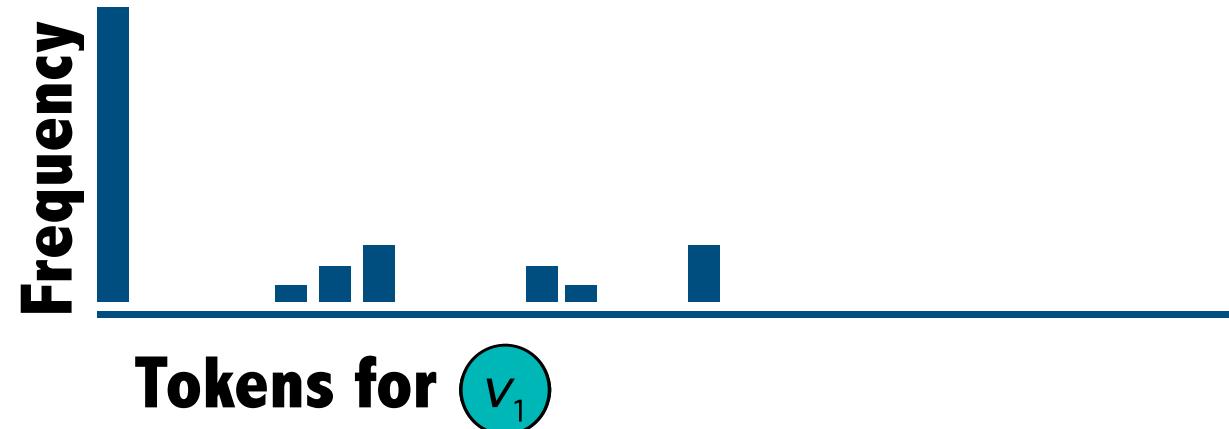


**Support Size Estimation
on Tokens**



$$\rightarrow \widehat{L}_0 \widehat{L}_1 = 350$$

**Choose Lengths that agree with
the Estimations**



$$\rightarrow \widehat{L}_1 \widehat{L}_2 = 1015$$

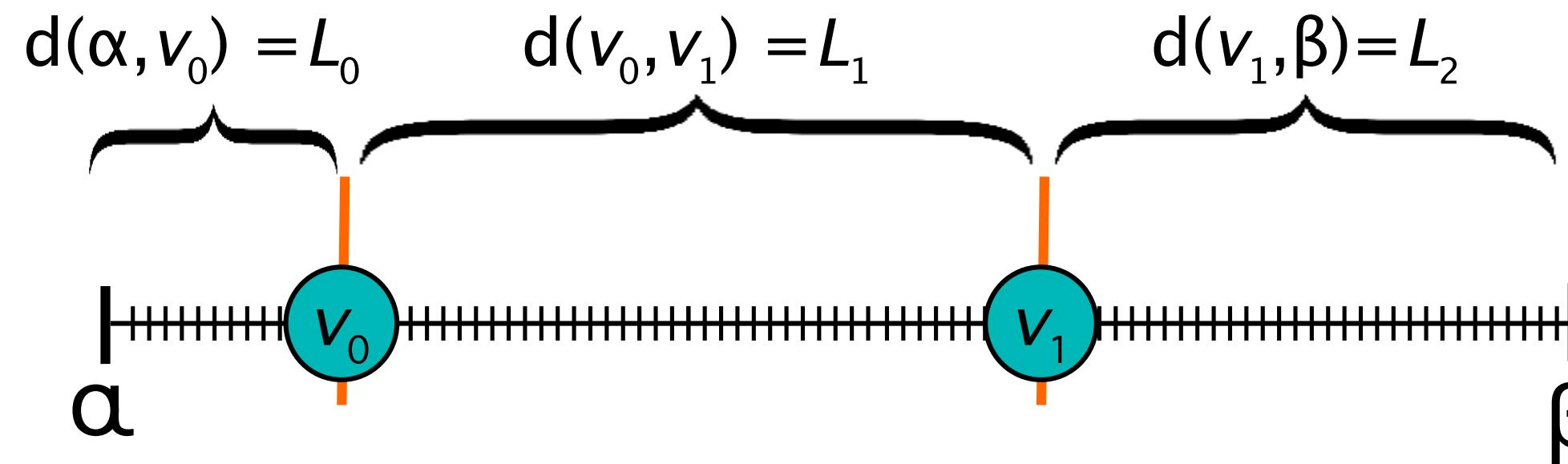


$$\rightarrow \widehat{L}_0 \widehat{L}_2 = 290$$

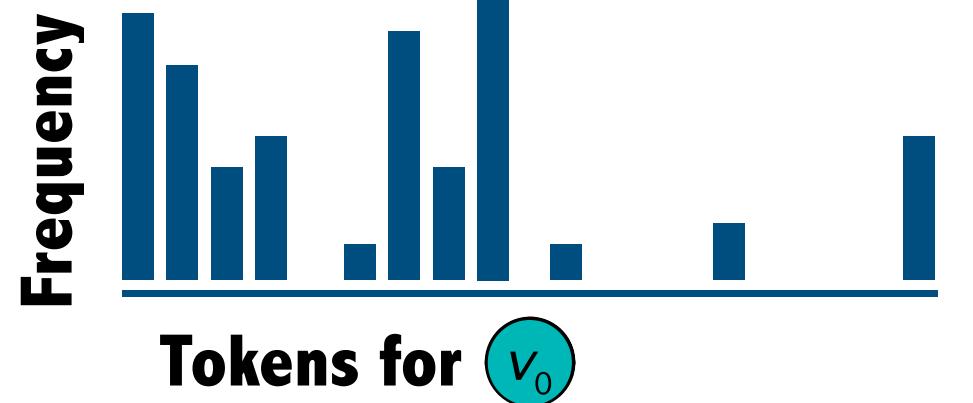


OVERVIEW OF THE ATTACK

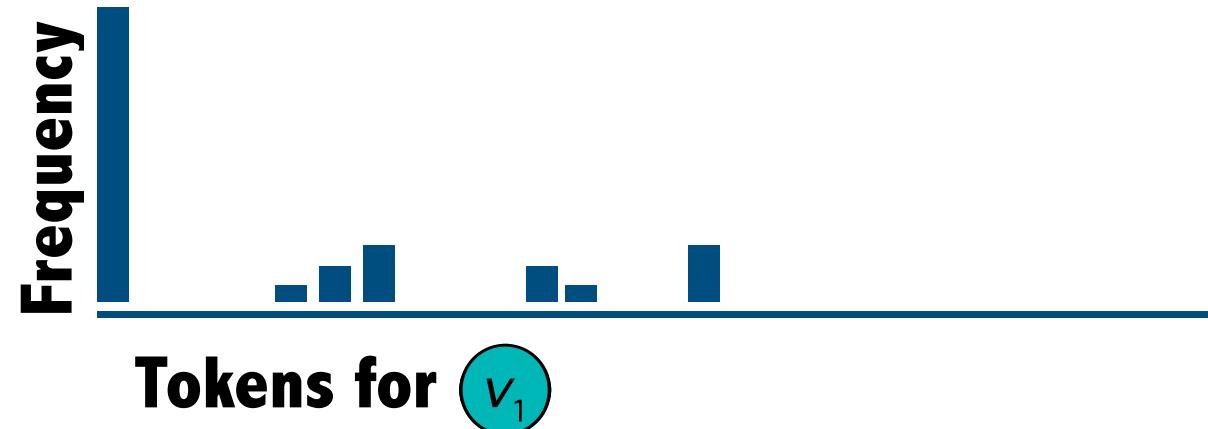
Plaintext:



**Support Size Estimation
on Tokens**



$$\widehat{L}_0 \widehat{L}_1 = 350$$



$$\widehat{L}_1 \widehat{L}_2 = 1015$$



$$\widehat{L}_0 \widehat{L}_2 = 290$$

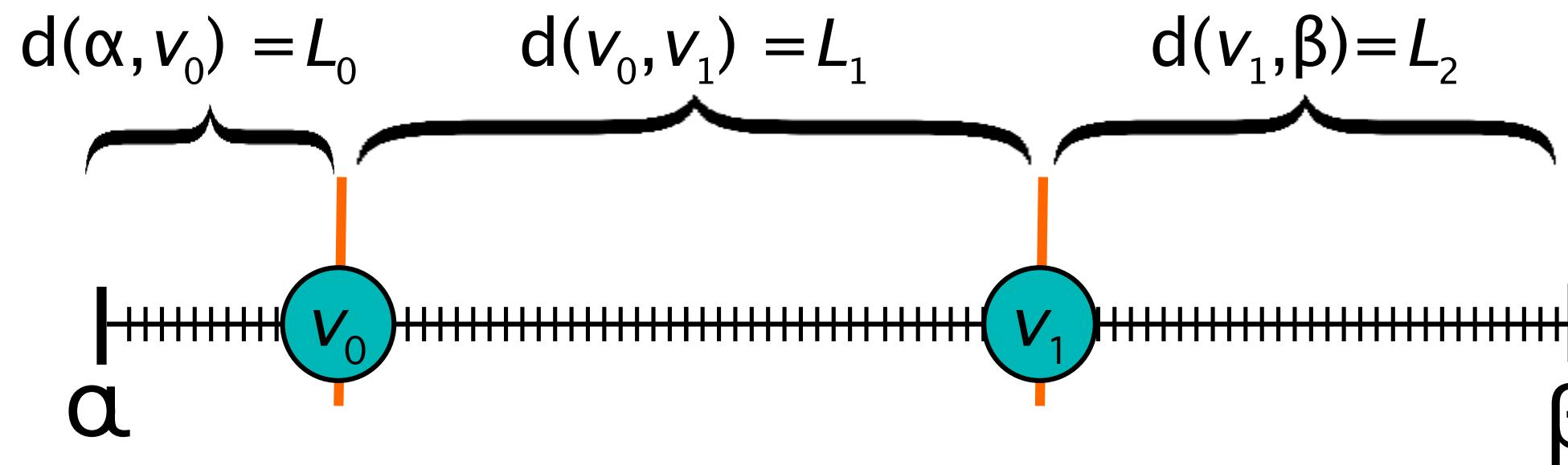
**Choose Lengths that agree with
the Estimations**

$$\begin{aligned} & \min_{L_0, L_1, L_2} ((L_0 \cdot L_1 - 350)^2 + (L_1 \cdot L_2 - 1015)^2 - (L_0 \cdot L_2 - 290)^2) \\ \text{s.t. } & \sum L_i = N \\ & L_i \geq 0 \end{aligned}$$



OVERVIEW OF THE ATTACK

Plaintext:



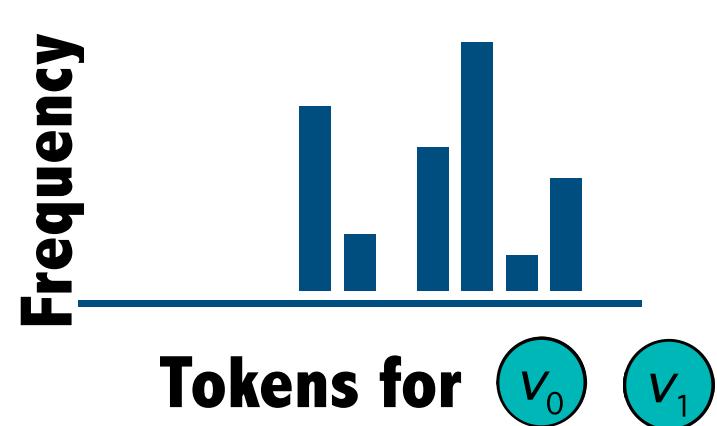
**Support Size Estimation
on Tokens**



$$\widehat{L}_0 \widehat{L}_1 = 350$$



$$\widehat{L}_1 \widehat{L}_2 = 1015$$



$$\widehat{L}_0 \widehat{L}_2 = 290$$

**Choose Lengths that agree with
the Estimations**

$$\min_{L_0, L_1, L_2} ((L_0 \cdot L_1 - 350)^2 + (L_1 \cdot L_2 - 1015)^2 - (L_0 \cdot L_2 - 290)^2)$$

s.t. $\sum L_i = N$

$$L_i \geq 0$$



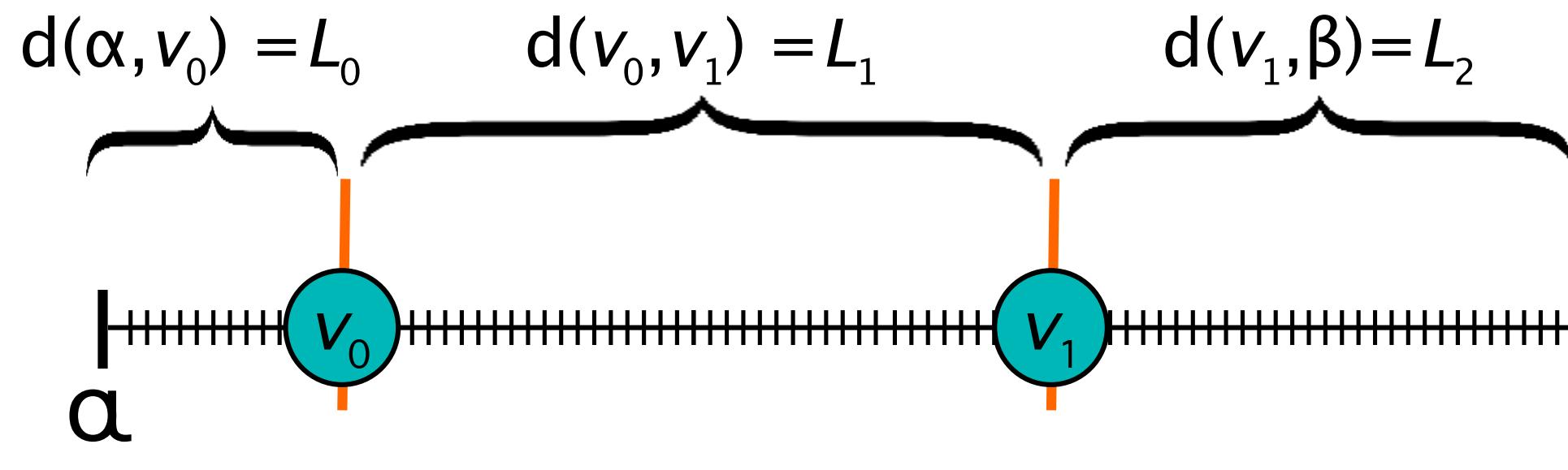
$$\min_{X_0, X_1, X_2} ((X_0 + X_1 - \log 350)^2 + (X_1 + X_2 - \log 1015)^2 - (X_0 + X_2 - \log 290)^2)$$

s.t. $\sum X_i = \log N$

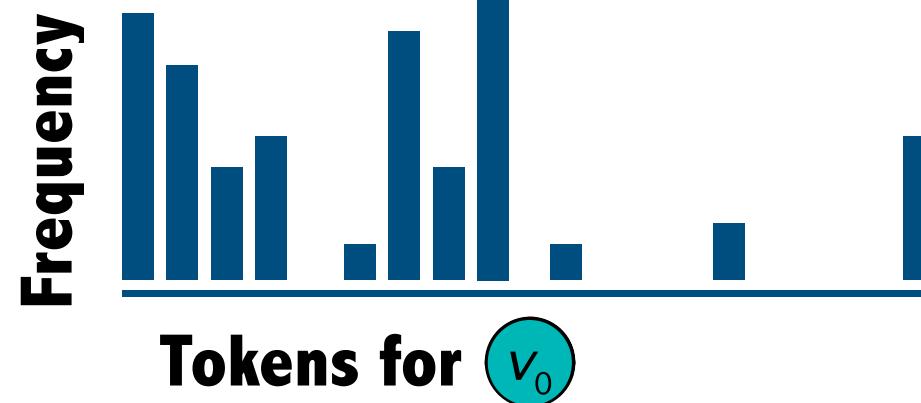


OVERVIEW OF THE ATTACK

Plaintext:



**Support Size Estimation
on Tokens**



$$\widehat{L}_0 \widehat{L}_1 = 350$$



$$\widehat{L}_1 \widehat{L}_2 = 1015$$



$$\widehat{L}_0 \widehat{L}_2 = 290$$

**Choose Lengths that agree with
the Estimations**

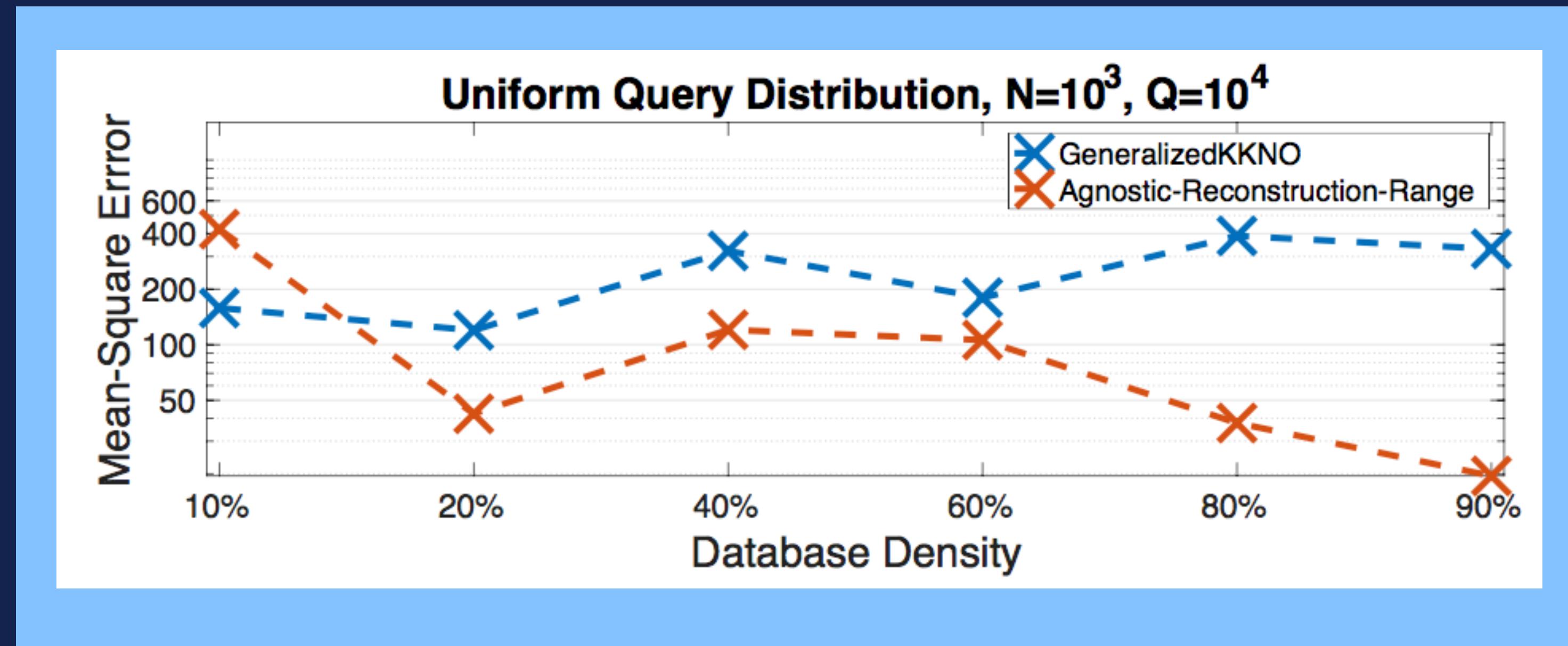
$$\begin{aligned} & \min_{L_0, L_1, L_2} ((L_0 \cdot L_1 - 350)^2 + (L_1 \cdot L_2 - 1015)^2 - (L_0 \cdot L_2 - 290)^2) \\ \text{s.t. } & \sum L_i = N \\ & L_i \geq 0 \end{aligned}$$

$$\begin{aligned} & \min_{X_0, X_1, X_2} ((X_0 + X_1 - \log 350)^2 + (X_1 + X_2 - \log 1015)^2 - (X_0 + X_2 - \log 290)^2) \\ \text{s.t. } & \sum X_i = \log N \end{aligned}$$



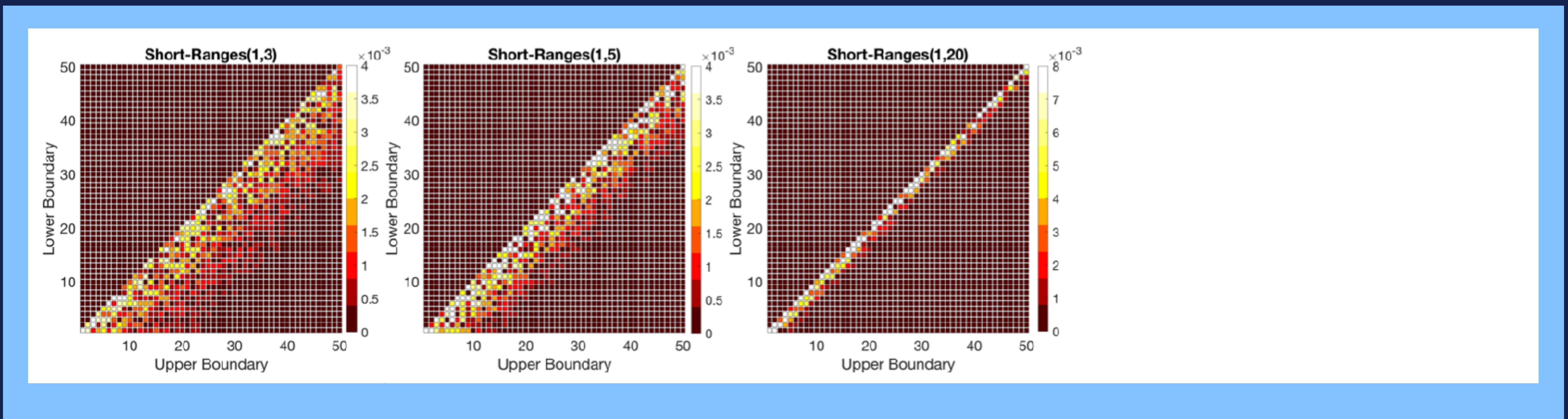
RANGE QUERIES

APPROXIMATE RECONSTRUCTION



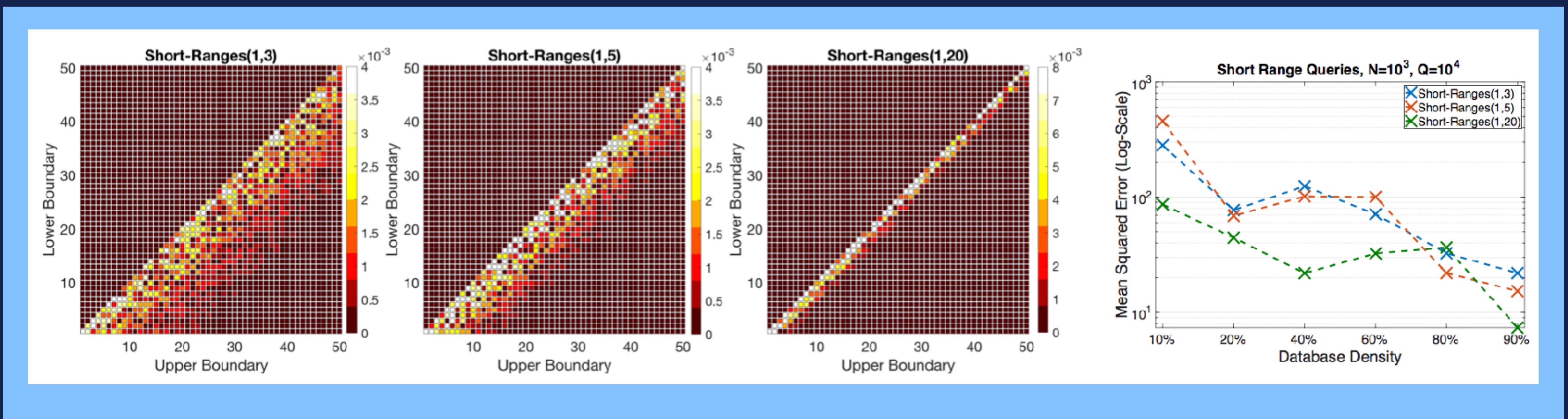


RANGE QUERIES APPROXIMATE RECONSTRUCTION





RANGE QUERIES APPROXIMATE RECONSTRUCTION





First attacks that combine Search Pattern and Access Pattern Leakage to **overcome strong assumptions** such as uniform query distribution.

S&P'20

The State of the Uniform: Attacks on Encrypted Databases Beyond the Uniform Query Distribution

Evgenios M. Kornaropoulos
UC Berkeley

Charalampos Papamanthou
University of Maryland

Roberto Tamassia
Brown University

Abstract—Recent foundational work on leakage-abuse attacks on encrypted databases has broadened our understanding of what an adversary can accomplish with a standard leakage profile. Nevertheless, all known value reconstruction attacks succeed under strong assumptions that may not hold in the real world. The most prevalent assumption is that queries are issued uniformly at random by the client. We present the first value reconstruction attacks that succeed *without any knowledge about the query or data distribution*. Our approach uses the search-pattern leakage, which exists in all known structured encryption schemes but has not been fully exploited so far. At the core of our method lies a support size estimator, a technique that utilizes the repetition of search tokens with the same response to estimate distances between encrypted values without any assumptions about the underlying distribution. We develop distribution-agnostic reconstruction attacks for both range queries and k -nearest-neighbor (k -NN) queries based on information extracted from the search-pattern leakage. Our new range attack follows a different algorithmic approach than state-of-the-art attacks, which are fine-tuned to succeed under the uniformly distributed queries. Instead, we reconstruct plaintext values under a variety of skewed query distributions and even outperform the accuracy of previous data distribution. In this paper, we take the next step

Value Reconstruction Attack Algorithms	Query Type	Assumptions				
		Query Distribution	Data Values in a Fixed Region	Known Data Distribution	Known Query Distribution	Dense Database
KPT	k -NN	Uniform	-	-	-	-
KKNO	Range	Uniform	-	-	-	-
LMP	Range	Agnostic	-	-	-	●
GLMP GENERALIZEDKKNO	Range	Uniform	-	-	-	-
GLMP APPROXVALUE	Range	Uniform	●	-	-	-
GLMP AOR to ADR	Range	Agnostic	-	●	●	-
This Work	k -NN & Range	Agnostic	-	-	-	-