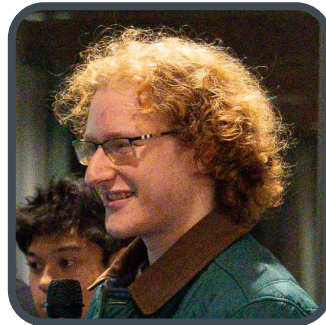


MEDUSA: Closing the Data Alignment Gap in Secure Analytics with Private Edit Distance



Adam Godel
agodel@bu.edu

Advisors:

Vasiliki Kalavri vkalavri@bu.edu	Mayank Varia varia@bu.edu	John Liagouris liagos@bu.edu
--	-------------------------------------	--

Performing Joint Analytics Requires Data Alignment

Data owner A

cust_name	zip_code	city
K. Ito	10012	NYC
A. Smith	02215	BOS

Secret table

cust_name	city	zip_code
K. Ito	NYC	10012
A. Smith	BOS	02215
J. Doe	BOS	02446

■ Attribute to compare
■ Attribute to impute

Schema matching

Imputation

Data owner B

custname	cty	zipcode
J. Doe	NULL	02446

Analytics systems based on secure multiparty computation (MPC)

CryptDough (SOSP'26)

ORQ (SOSP'25)

CoVault (USENIX Security'25)

Piranha (USENIX Security'22)

...all assume aligned inputs

The Data Alignment Gap

Real data is incomplete and misaligned



Aligning it relies on edit distance



MEDUSA facilitates end-to-end secure data alignment using a state-of-the-art MPC edit distance primitive

Edit Distance

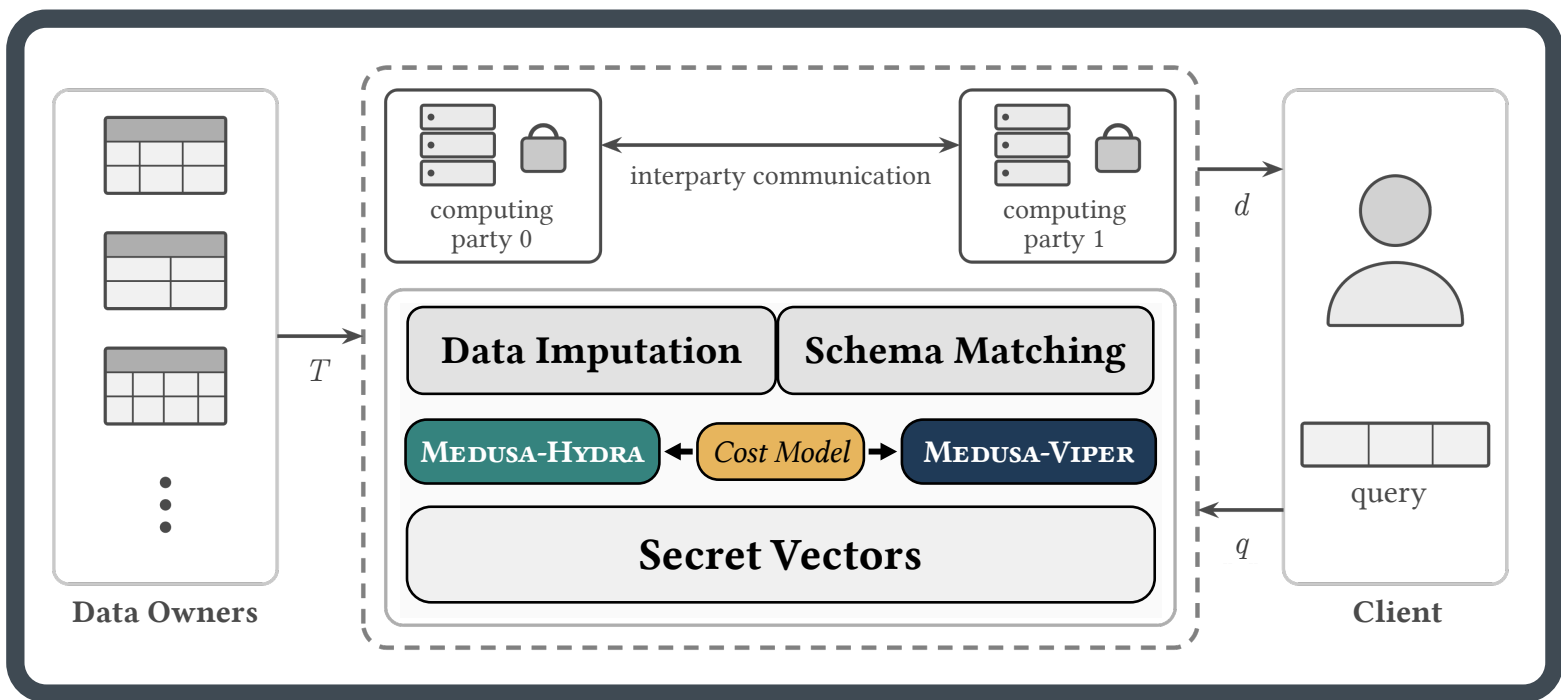
p	h	o	n	e	n	u	m
p	h	o	n	e	-	n	o

One insertion, one deletion, one substitution:

distance 3

Data owners may not want to reveal their raw data, or even their schemas.

MEDUSA: Private Edit Distance on Secret Vectors



Two private edit distance primitives:

MEDUSA-VIPER

Yao sharing: one string vs. one string in a constant number of rounds

MEDUSA-HYDRA

Boolean sharing: one string vs. every record of a table, in parallel

Our starting point is Myers (JACM'99): $O(mn) \rightarrow O(n)$ communication rounds

1 MPC-Friendly Myers

Secret-index lookups $\rightarrow$ one SIMD equality round; stacking independent steps halves rounds per iteration.

2 Record-Parallel Myers

The table is transposed so every record advances through each Myers iteration in lockstep.

3 Cost-Based Switching

An analytic cost model picks MEDUSA-HYDRA or batched MEDUSA-VIPER per workload from input size and network.

4 Engine Integration

A vectorized, pipelined garbling backend on CryptDough (SOSP'26), which had no Yao sharing support, to implement MEDUSA-VIPER.

End-to-End Private Data Alignment

Private Data Imputation

Goal: Select a uniformly random value under some private threshold.

Insight: Obviously select it by picking the first eligible value after a shuffle.

d	2	6	7	3	5
$d \leq 3$	1	0	0	1	0
<i>Shuffle</i>	0	1	0	1	0
$\neg$ PrefixOR	1	1	0	0	0
<i>sel</i>	0	1	0	0	0

1 Edit distance to every record

2 Compare with private threshold

3 Obviously shuffle

4 Prefix-OR picks the first eligible

Private Schema Matching

Goal: We build off prior work on MPC stable matching to create a protocol to stable match edit distances.

Insight: Since the preferences are edit distances, they will be symmetric, so we can just run two oblivious argmin operations.

	R_0	R_1	R_2	R_3	
S_0	2	5	7	9	
S_1	3	6	8	4	
S_2	1	6	2	7	
S_3	8	4	5	3	

■ Accepted
■ Rejected

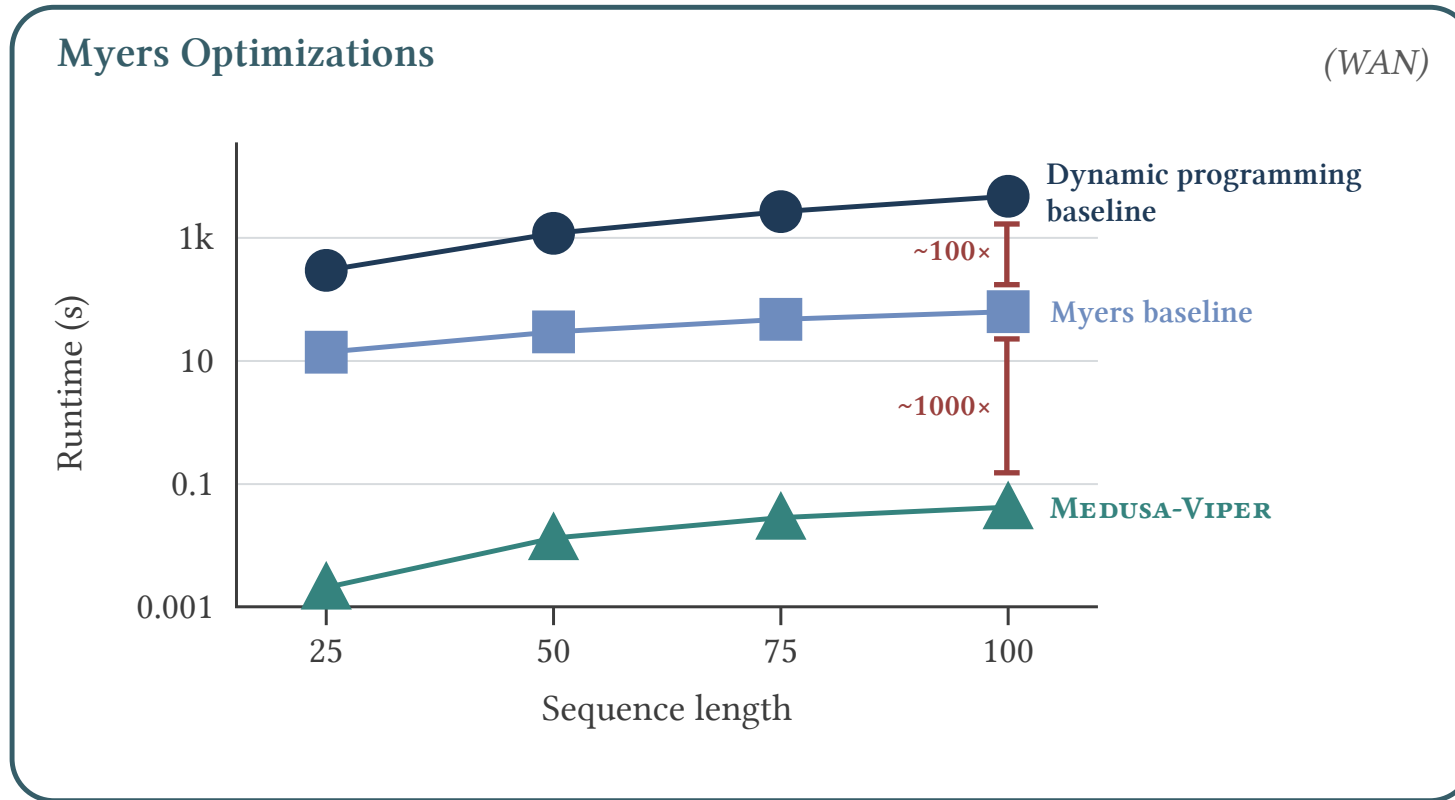
1 All pairwise distances

2 Suitors propose to the nearest

3 Reviewers accept the closest

4 Repeat until no rejections

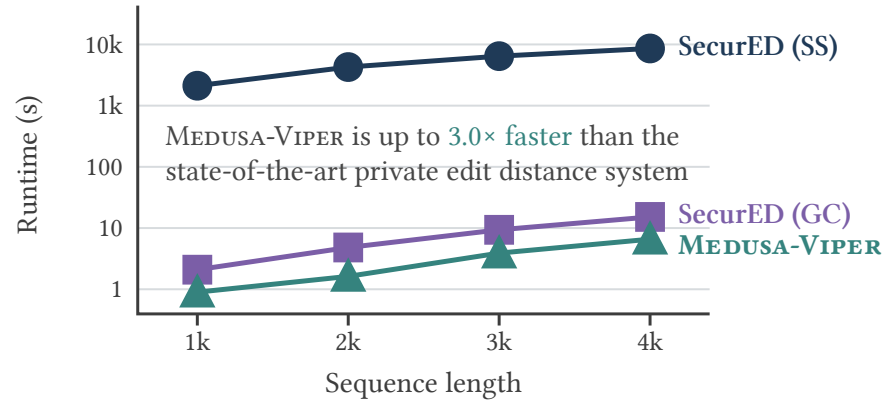
MEDUSA Outperforms the State-of-the-Art Systems



MEDUSA Outperforms the State-of-the-Art Systems

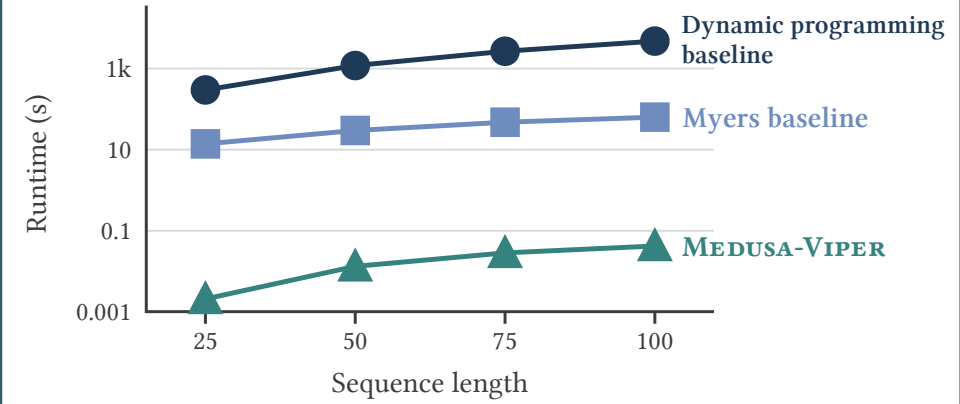
Comparison with SecurED

(WAN)



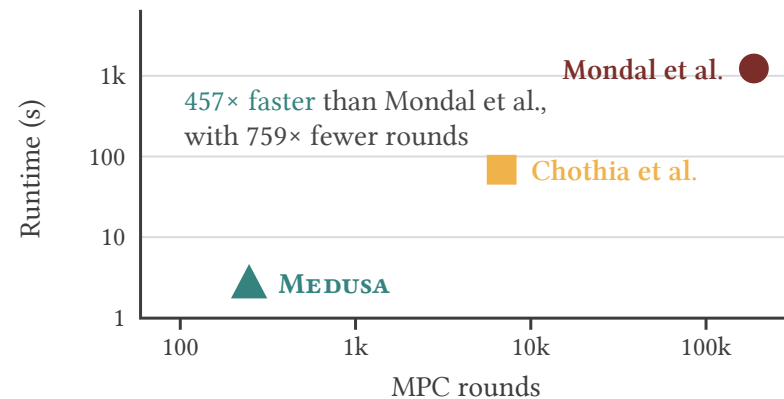
Myers Optimizations

(WAN)

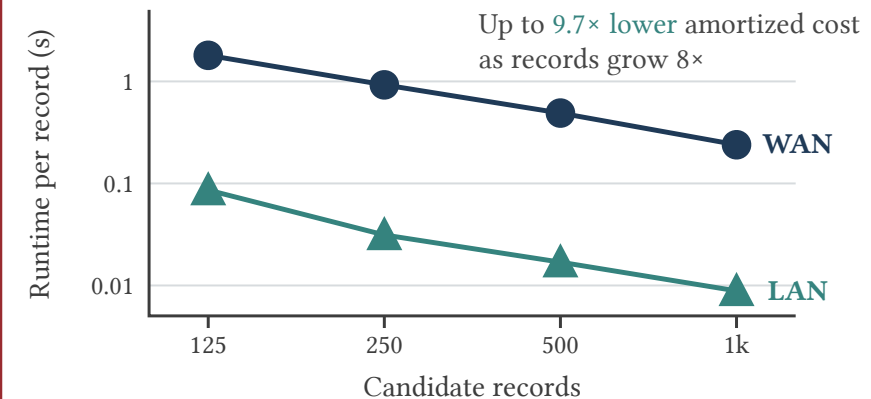


Private Schema Matching

(WAN)



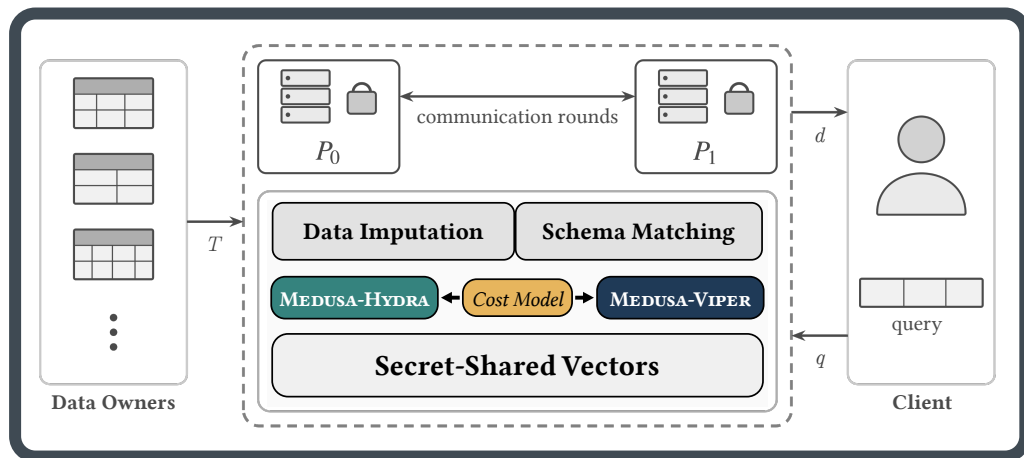
Private Data Imputation



MEDUSA: Closing the Data Alignment Gap in Secure Analytics with Private Edit Distance

Adam Godel

agodel@bu.edu



Fast Private Edit Distance

MPC-friendly, record-parallel
Myers, up to $3.0\times$ faster than
SecurED (PETS'25)

Cost-Based Mode Switching

An analytic cost model picks
MEDUSA-HYDRA or batched
MEDUSA-VIPER per workload

End-to-End Data Alignment

First private imputation &
schema matching based on
edit distance

Engine Integration

A vectorized garbling
pipeline integrated on
CryptDough (SOSP'26)