CMPUT 626 - A2
Machine Learning and Practical Privacy

Thinking About Cryptography 2

Fall 2023, Tuesday/Thursday 3:30-4:50pm

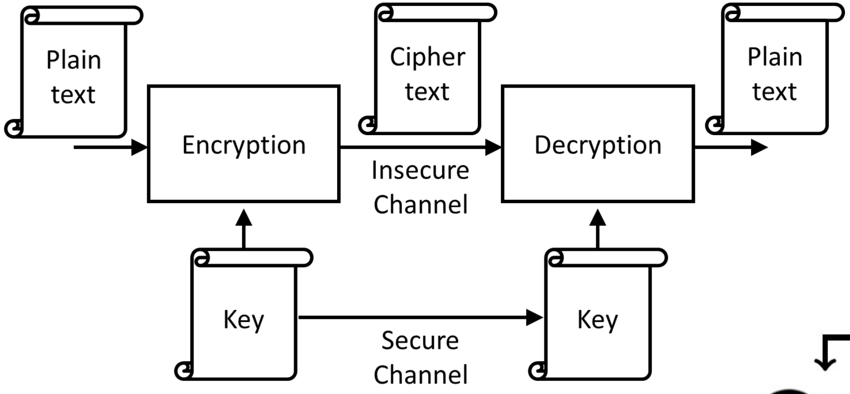
Admin Stuff

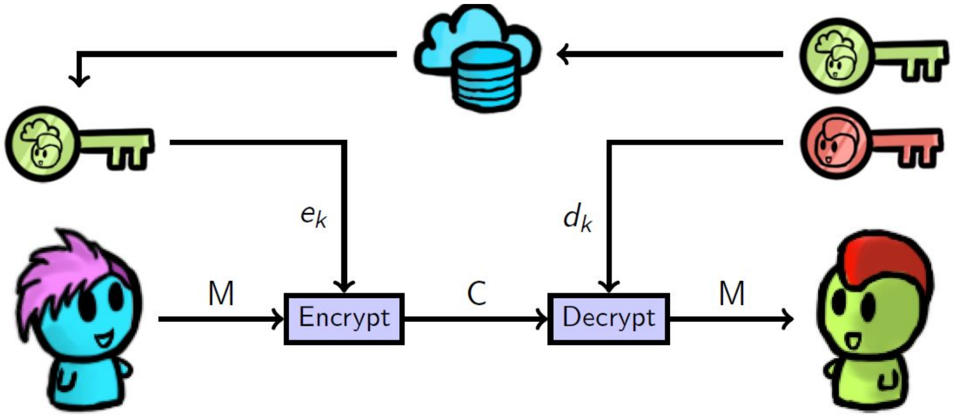
Send me your paper preferences TODAY.

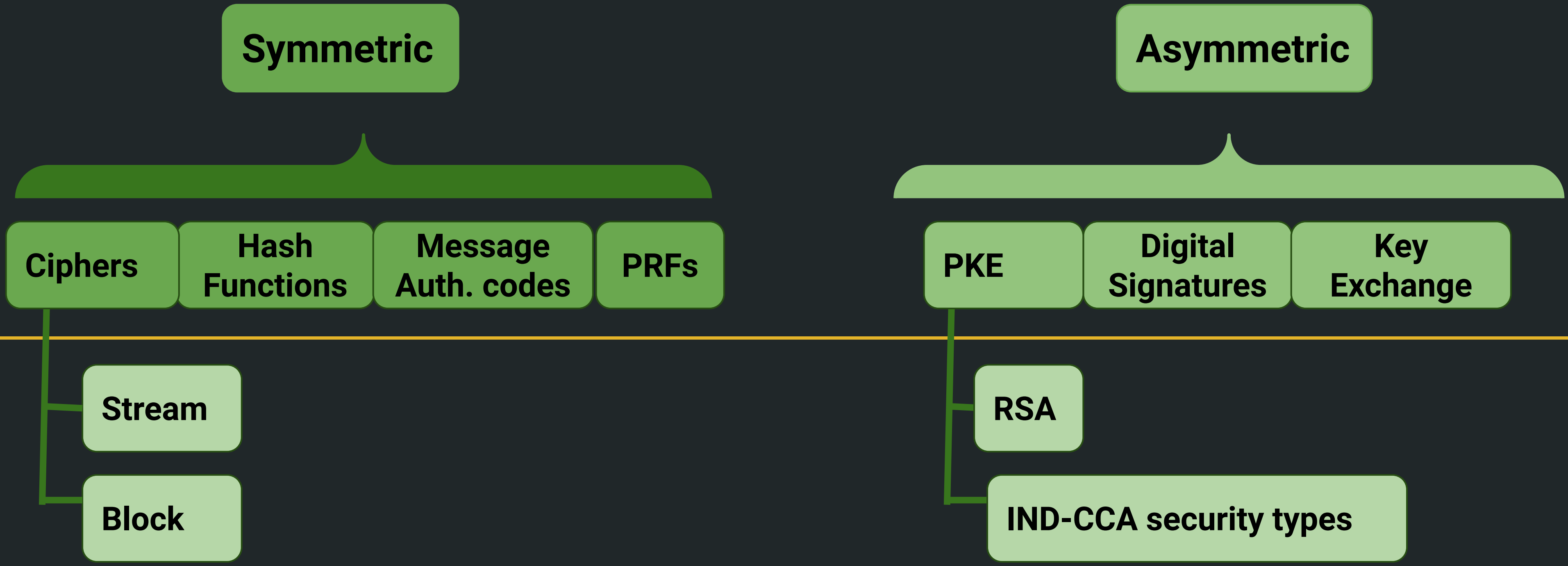
Include any scheduling conflicts.

If you just joined, by TOMORROW NOON.

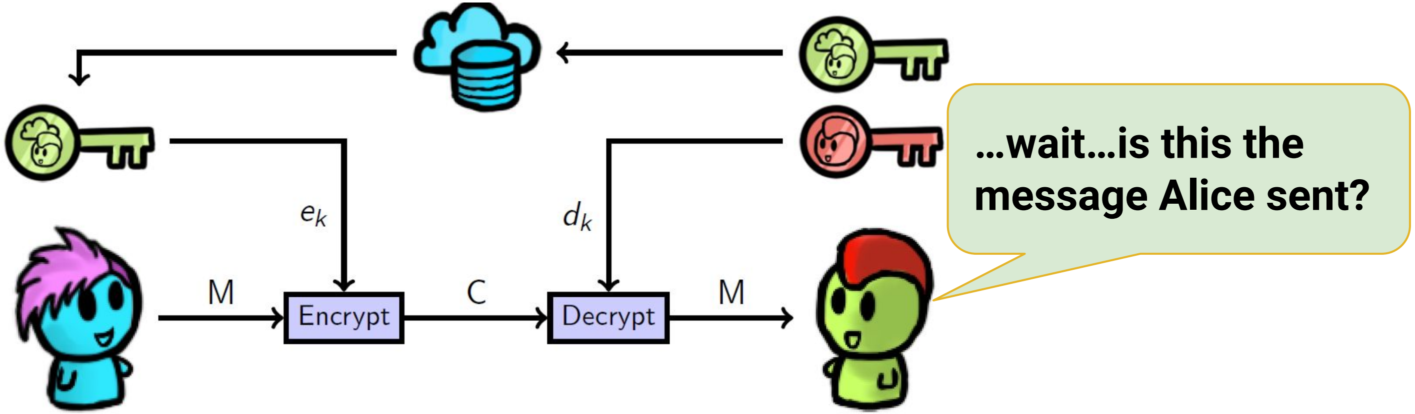
Block/Stream Ciphers, Public Key Cryptography…



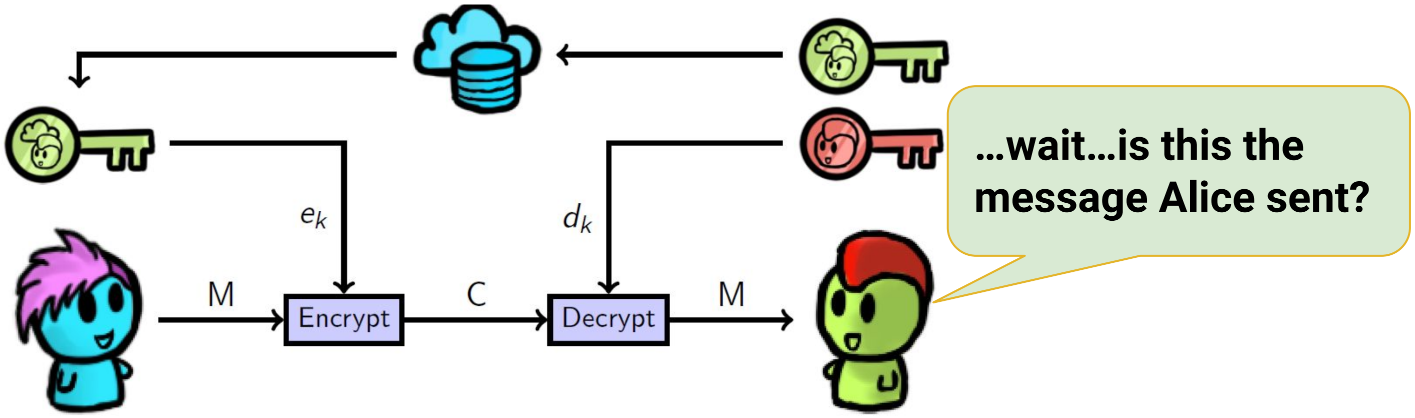



Symmetric
Asymmetric
Ciphers
Hash Functions
Message Auth. codes
PRFs
PKE
Digital Signatures
Key Exchange
Stream
Block
RSA
IND-CCA security types

Detect? Messages Changed in Transit



Detect? Messages Changed in Transit

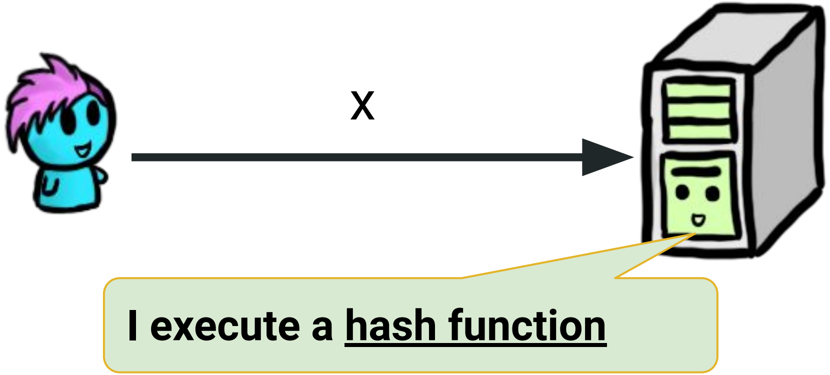


Checksums, appended so Bob can verify it

Not. Good. Enough.

Goal: Make it hard for Mallory to find a second message with the same checksum as the "real" one

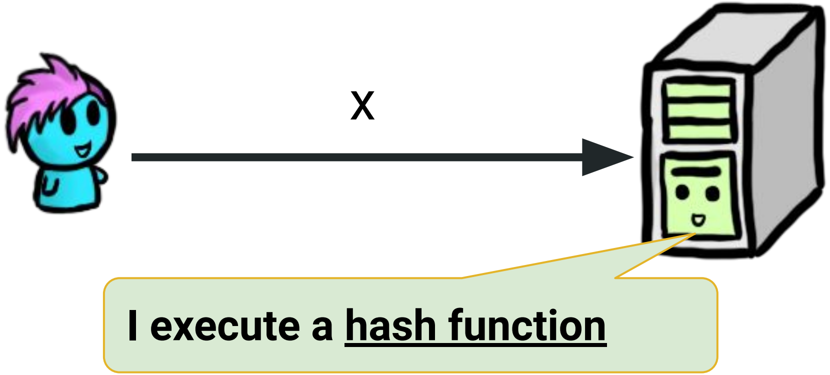
Towards Integrity: Cryptographic Hash Functions



Common examples:

- MD5, SHA-1, SHA-2, SHA-3 (aka Keccak after 2012)

Towards Integrity: Cryptographic Hash Functions

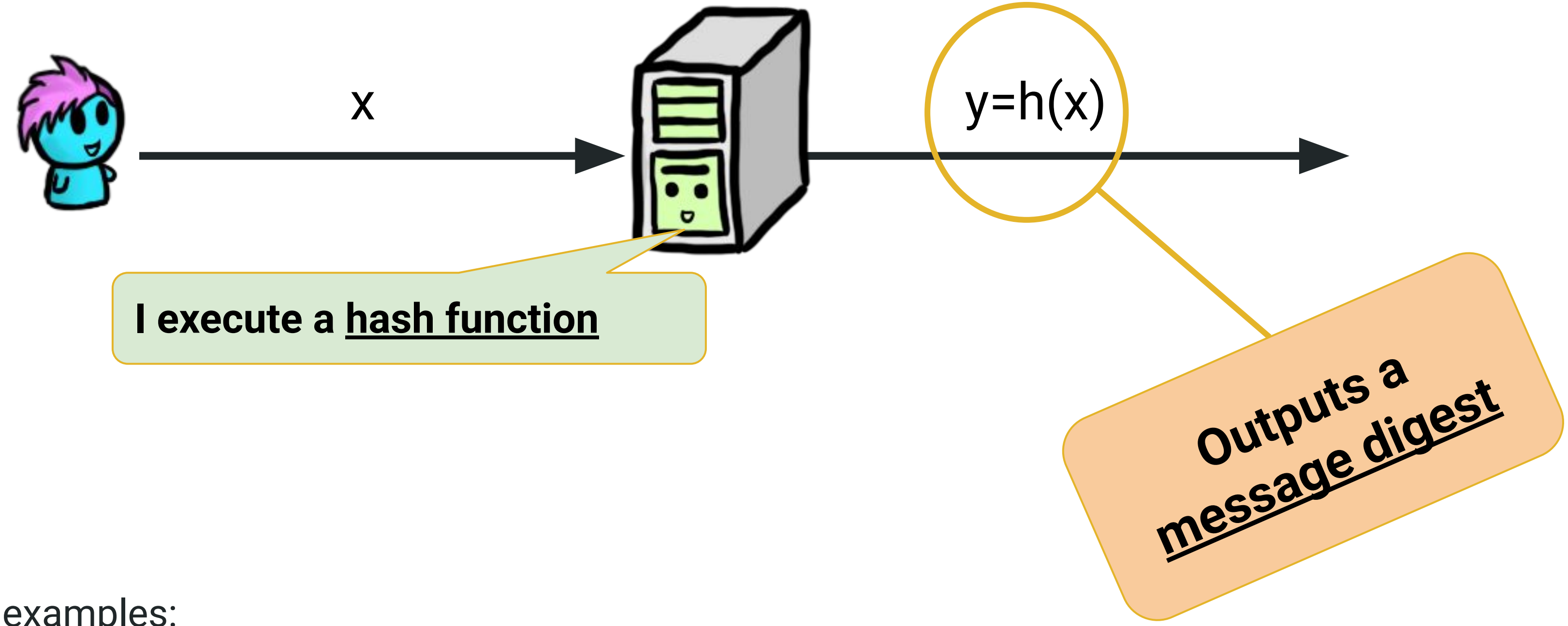


Takes an arbitrary length string, and computes a fixed length string.

Common examples:

- MD5, SHA-1, SHA-2, SHA-3 (aka Keccak after 2012)

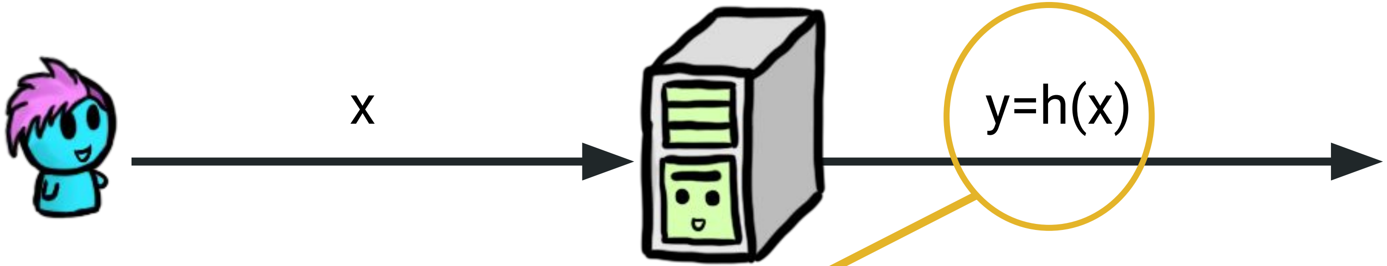
Towards Integrity: Cryptographic Hash Functions



Common examples:

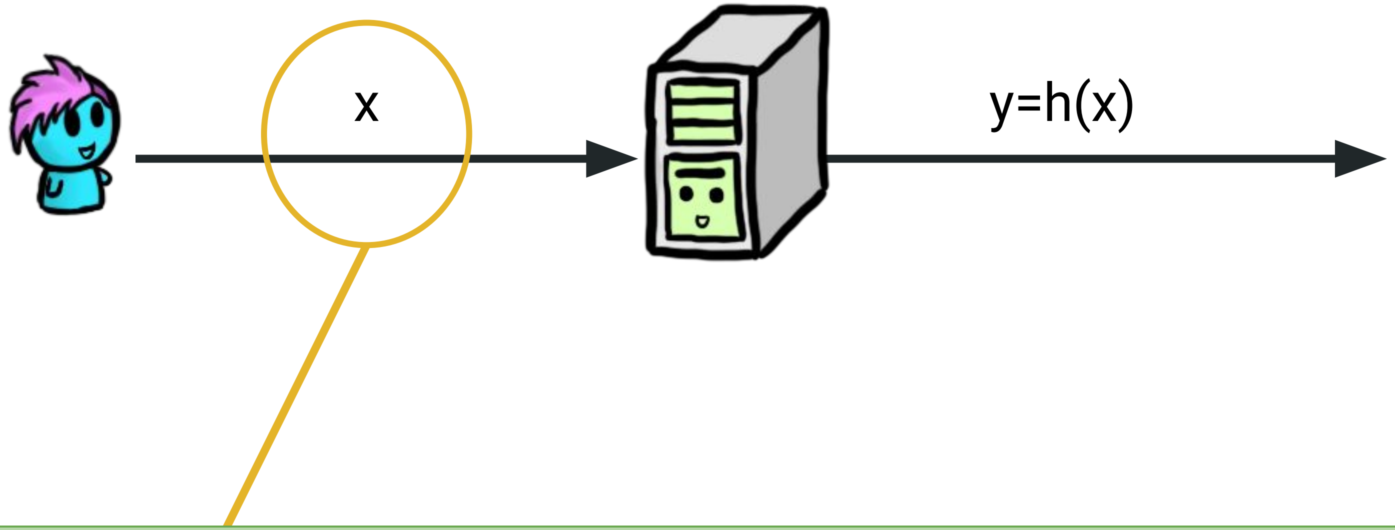
- MD5, SHA-1, SHA-2, SHA-3 (aka Keccak after 2012)

Properties: Preimage-Resistance



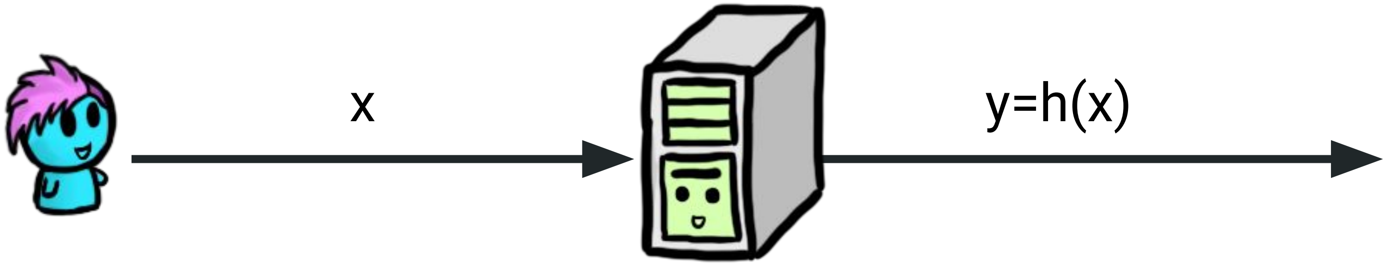
Goal: Given y, “hard” to find x such that h(x) = y

Properties: Second Preimage-Resistance



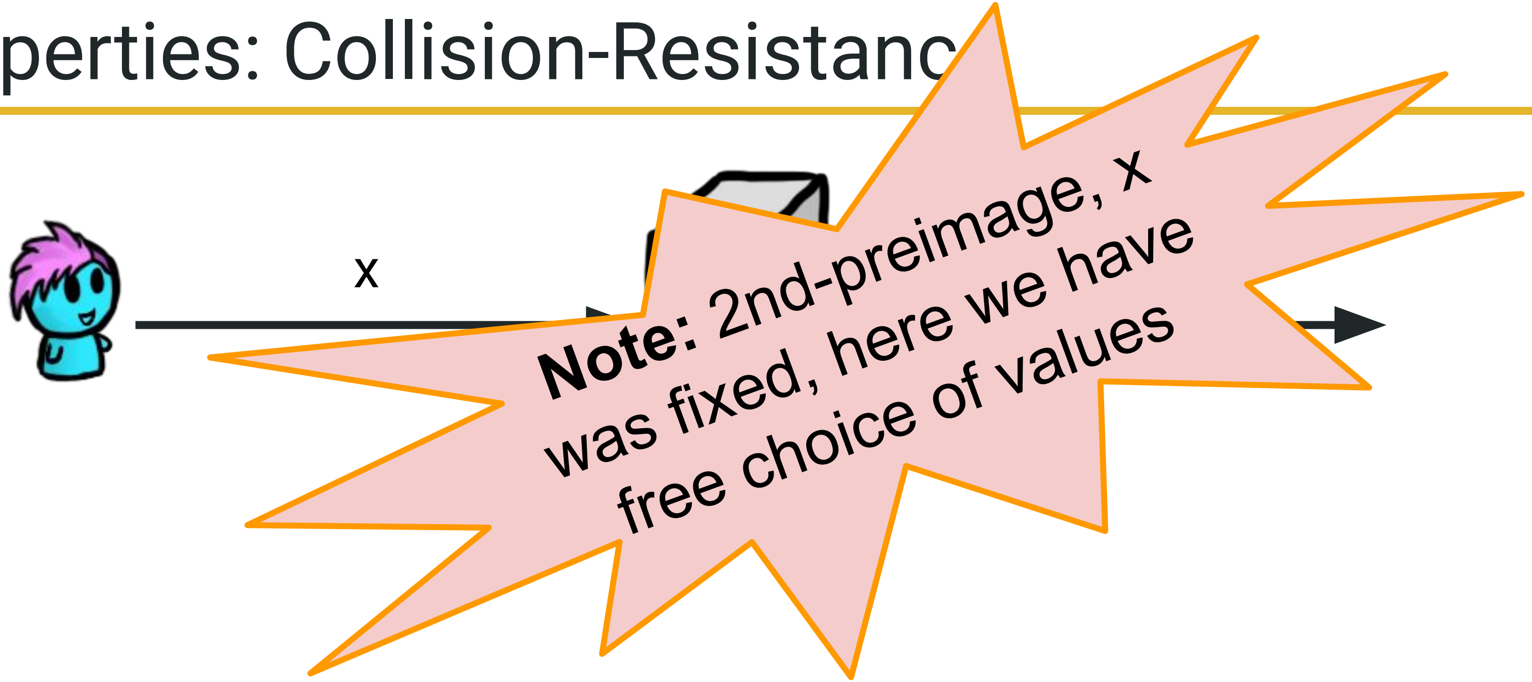
Goal: Given x, “hard” to find x’ <> x such that h(x) = h(x’)

Properties: Collision-Resistance



Goal: It’s hard to find any two distinct x, x’ such that h(x) = h(x’)

Properties: Collision-Resistanc



Goal: It’s hard to find any two distinct x, x’ such that h(x) = h(x’)

Making it too hard to break these properties?

- SHA-1: takes 2^{160} work to find a preimage or second image
- SHA-1: takes 2^{80} to find a collision using brute-force search

Making it too hard to break these properties?

- SHA-1: takes 2^{160} work to find a preimage or second image
- SHA-1: takes 2^{80} to find a collision using brute-force search

There are faster ways to find collisions in SHA-1 or MD5

Collisions and the Birthday Paradox

Collisions are easier due to the birthday paradox

Collisions and the Birthday Paradox

Collisions are easier due to the birthday paradox

What's the probability two of us have the same birthday?

Collisions and the Birthday Paradox

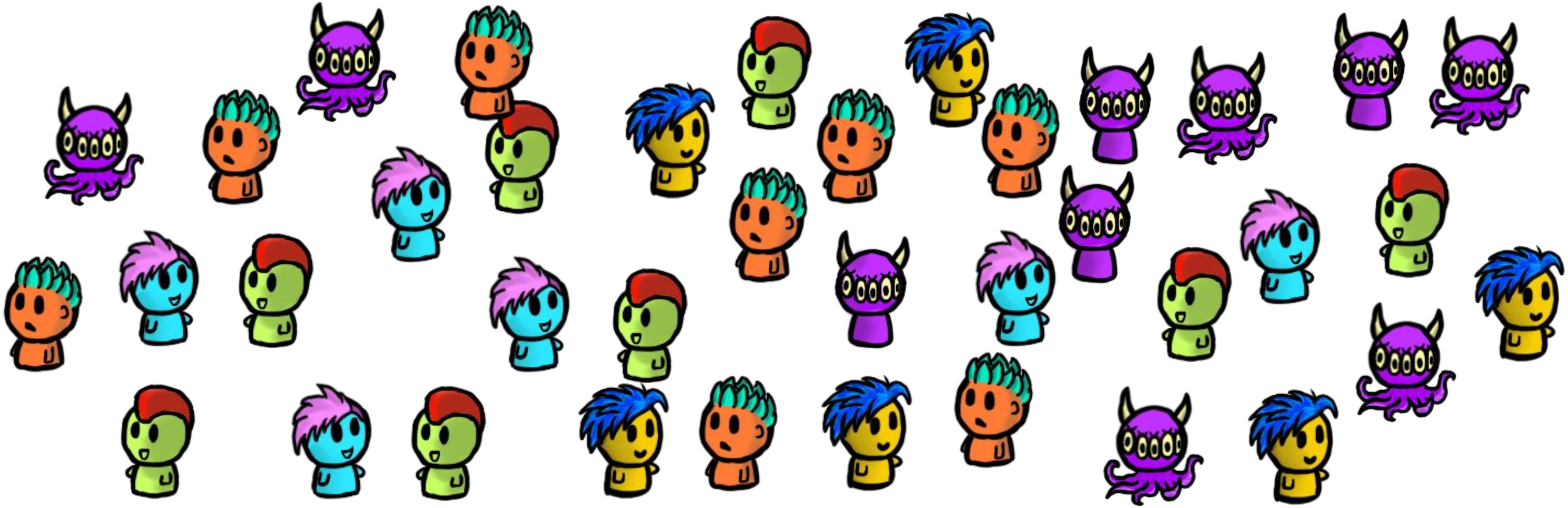
Collisions are easier due to the birthday paradox

What's the probability two of us have the same birthday?

There's 23 of us, so larger than 50%!!

Collisions and the Birthday Paradox

Collisions are easier due to the birthday paradox



Collisions and the Birthday Paradox

Collisions are easier due to the birthday paradox

Collisions and the Birthday Paradox

Collisions are easier due to the birthday paradox

Collisions and the Birthday Paradox

Collisions are easier due to the birth... pa...

There's ...

Not the end of our problems

How about a bad example? (Integrity over Conf.)

Q: What can Mallory do to send the message she wants (change it)?

How about a bad example? (Integrity over Conf.)

Q: What can Mallory do to send the message she wants (change it)?

A: Just change it…Mallory can compute the new hash herself.

How about a less bad example? (Integrity & Conf.)

Q: What can Mallory do to send the message she wants (change it)?

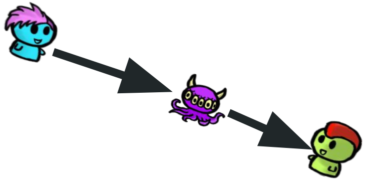
How about a less bad example? (Integrity & Conf.)

Q: What can Mallory do to send the message she wants (change it)?

A: Still. Just change it.

Limitations for Cryptographic Hash Functions

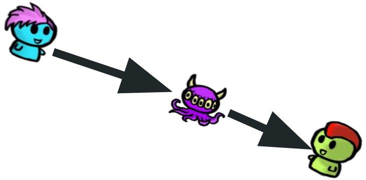
- Integrity guarantees only when there is a <u>**secure**</u> way of sending/storing the message digest

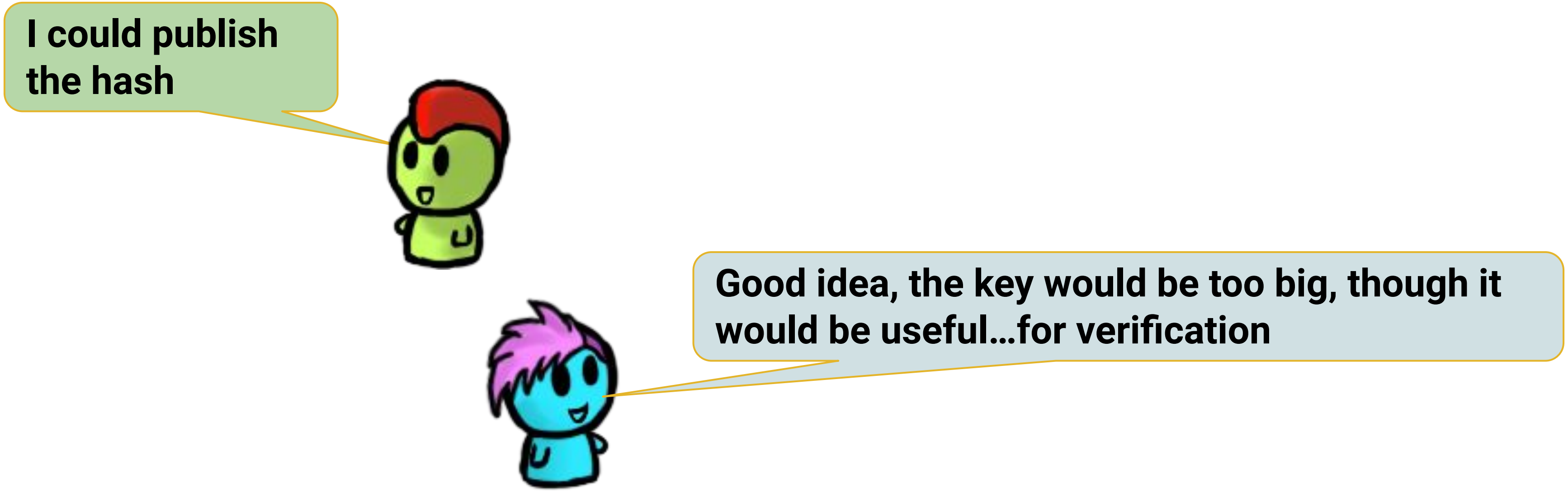


I could publish the hash

Limitations for Cryptographic Hash Functions

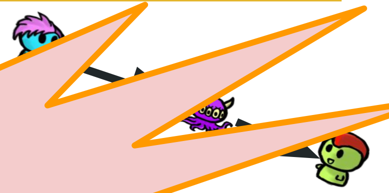
- Integrity guarantees only when there is a <u>**secure**</u> way of sending/storing the message digest





Limitations for Cryptographic Hash Functions

- Integrity guarantees only when there is a ecure way of sending/storing the message d



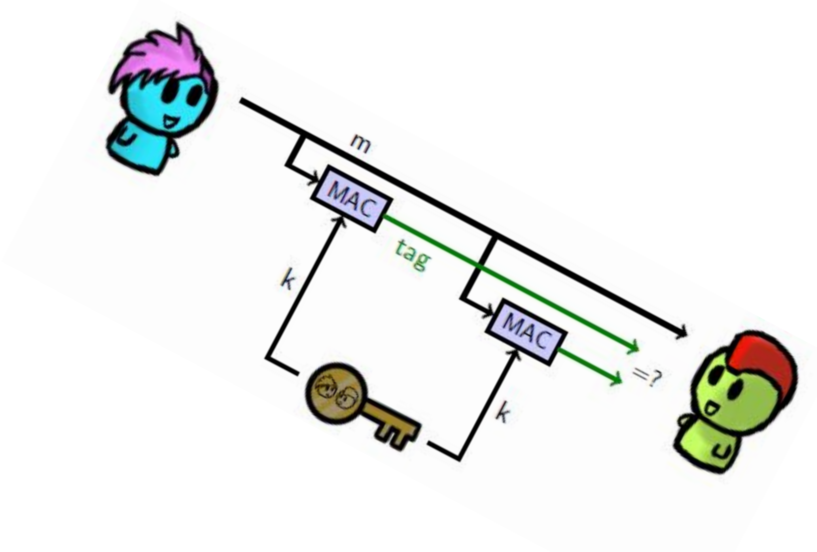
I could publish the hash

What if…we don't have an external channel?

idea, ey would be too big, though it
ould be use ul…for verification

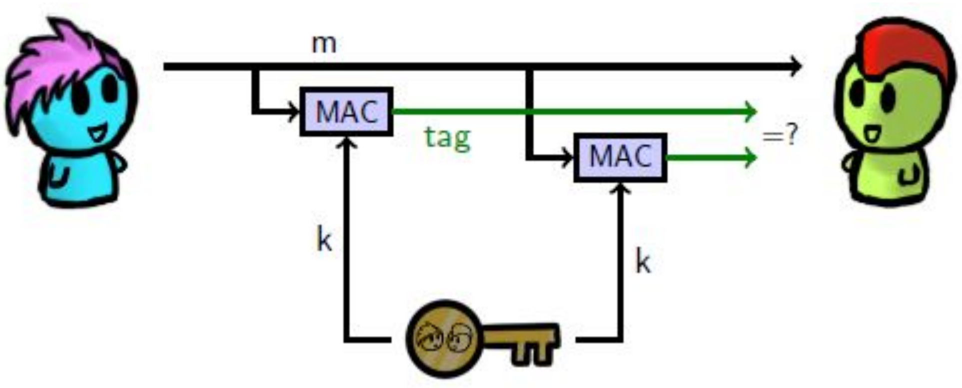
Authentication and Hash Functions

- Use “keyed hash functions”
- Requires the key to generate or check the hash value (tag)



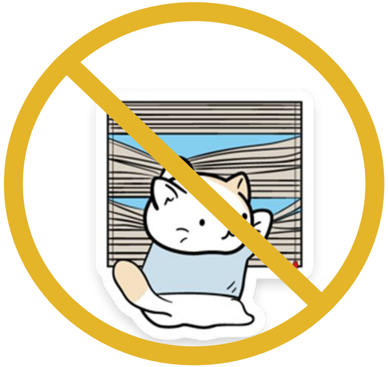
Called: Message authentication codes (MACs)

Message Authentication Codes (MACs)



Use “keyed hash functions”
e.g., SHA-1-HMAC, SHA-256-HMAC, CBC-MAC

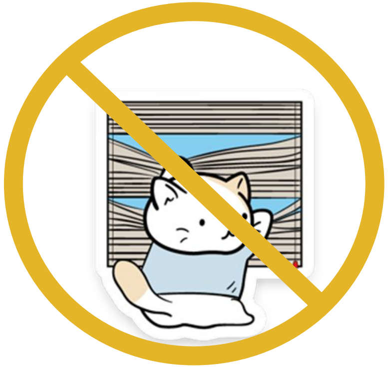
Combine Ciphers and MACs



Confidentiality

Integrity

Combine Ciphers and MACs



Confidentiality

Integrity

Practical systems need both

But how to combine them?

- MAC-then-Encrypt versus
- Encrypt-and-MAC versus
- Encrypt-then-MAC

But how to combine them?

- MAC-then-Encrypt versus
- Encrypt-and-MAC versus
- Encrypt-then-MAC

Ideally, there is an authenticated encryption mode that combines them…but…

Examples that do:
- GCM, CCM, or OCB

Make it work?

- Alice and Bob have a secret key **k** for a cryptosystem
- Also, a secret key **K'** for their MAC

Consider: How can Alice build a message for Bob in the following three scenarios.

MAC-then-Encrypt

- Alice and Bob have a secret key **k** for a cryptosystem and a secret key **K’** for their MAC
- Compute the MAC on the message, then encrypt the message and MAC together, and send that ciphertext.

Encrypt-and-MAC:

- Alice and Bob have a secret key **k** for a cryptosystem and a secret key **K’** for their MAC
- Compute the MAC on the message, the encryption of the message, and send both.

Encrypt-then-MAC:

- Alice and Bob have a secret key **k** for a cryptosystem and a secret key **K’** for their MAC
- Encrypt the message, compute the MAC on the encryption, send encrypted message and MAC

Which order is correct?

Usually: we want the receiver to verify the MAC first!

Q: Which should be recommended then?

$E_k(m||MAC_{K'}(m))$ vs. $E_k(m)||MAC_{K'}(m)$ vs. $E_k(m)||MAC_{K'}(E_k(m))$

Which order is correct?

Usually: we want the receiver to verify the MAC first!

Q: Which should be recommended then?

$E_k(m||MAC_{K'}(m))$ vs. $E_k(m)||MAC_{K'}(m)$ vs. $E_k(m)||MAC_{K'}(E_k(m))$

Recommended: Encrypt-then-MAC, $E_k(m)||MAC_{K'}(E_k(m))$

Which order is correct?

Usually: we want the receiver to verify the MAC first!

Q: Which should be recommended then?

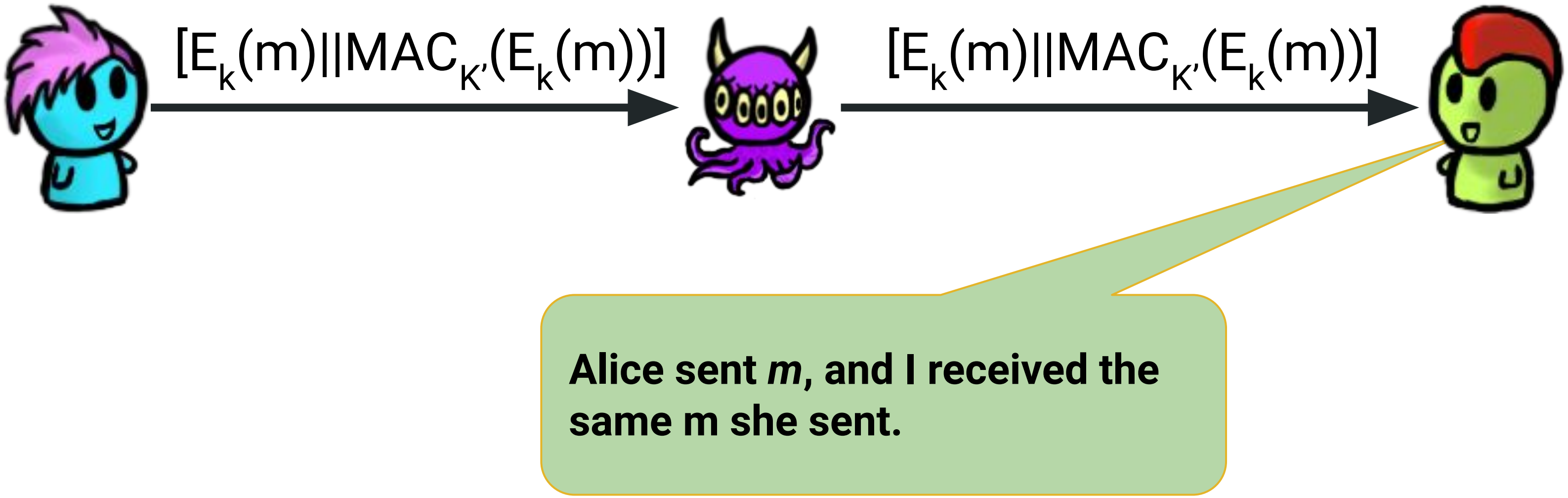
$E_k(m||MAC_{K'}(m))$ vs. $E_k(m)||MAC_{K'}(m)$ vs. $E_k(m)||MAC_{K'}(E_k(m))$

Recommended: Encrypt-then-MAC, $E_k(m)||MAC_{K'}(E_k(m))$

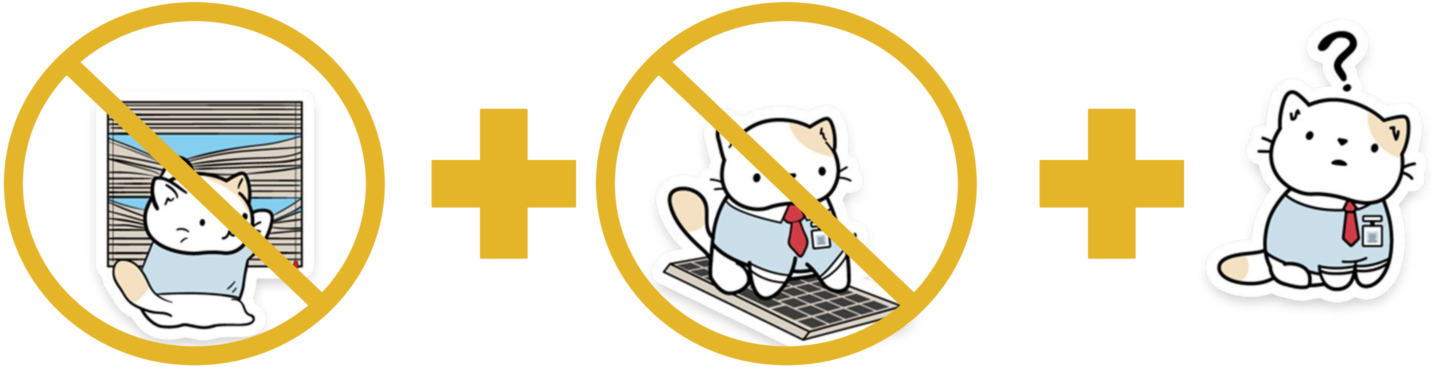
Why though?

More properties that matter?

Repudiation



Repudiation



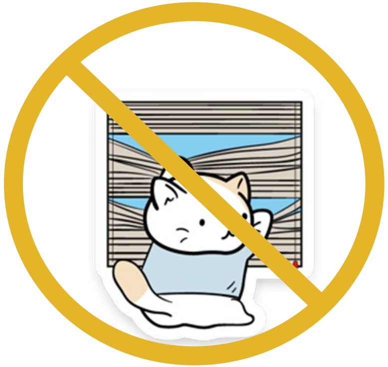
Confidentiality **Integrity** **Authentication**

Repudiation

$[E_k(m)||MAC_{K'}(E_k(m))]$

$[E_k(m)||MAC_{K'}(E_k(m))]$

Almost, but not quite a signature



Confidentiality + **Integrity** + **Authentication**

Repudiation

Almost, but not quite a signature…So…you're saying Bob can't prove Alice sent m?

Repudiation

Almost, but not quite a signature…So…you're saying Bob can't prove Alice sent m?

Q: Why can't Bob prove it?

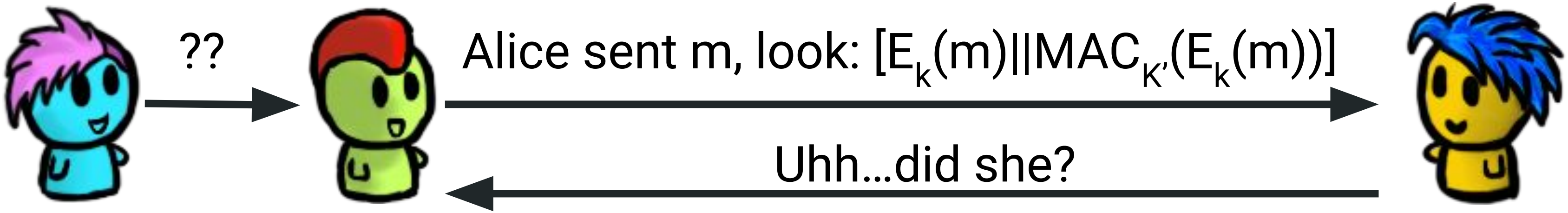
Repudiation

Almost, but not quite a signature…So…you're saying Bob can't prove Alice sent m?

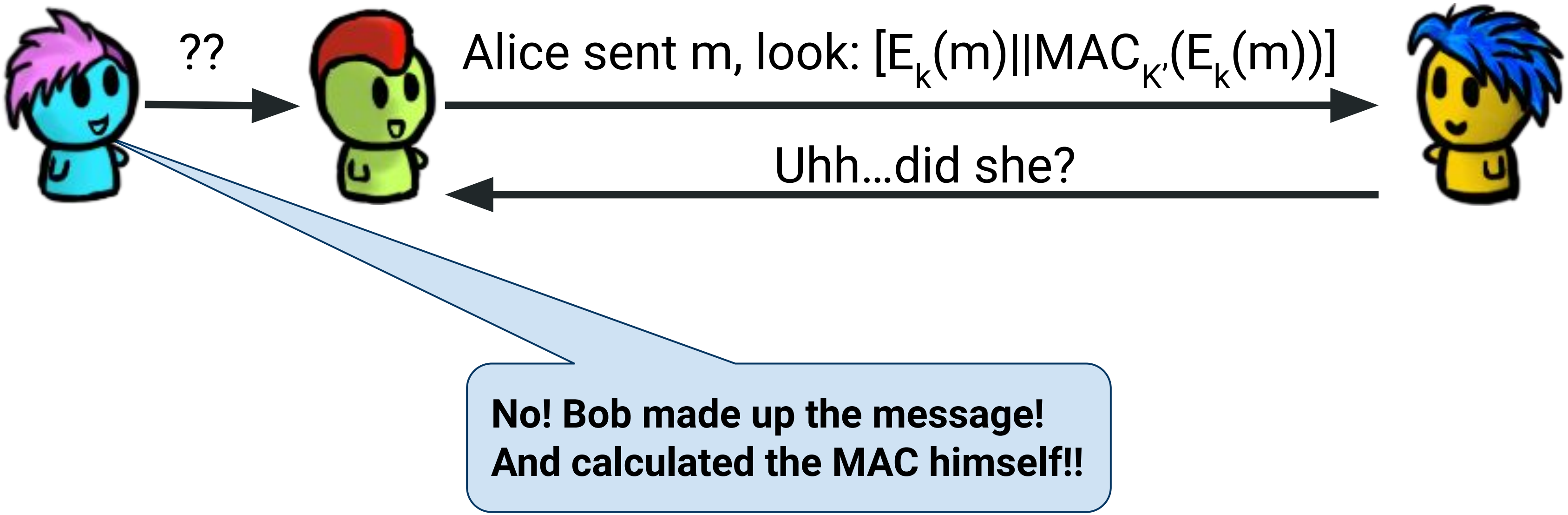
Q: Why can't Bob prove it?

A: Either Alice or Bob could create any message and MAC combo…also Carol doesn't know the secret keys.

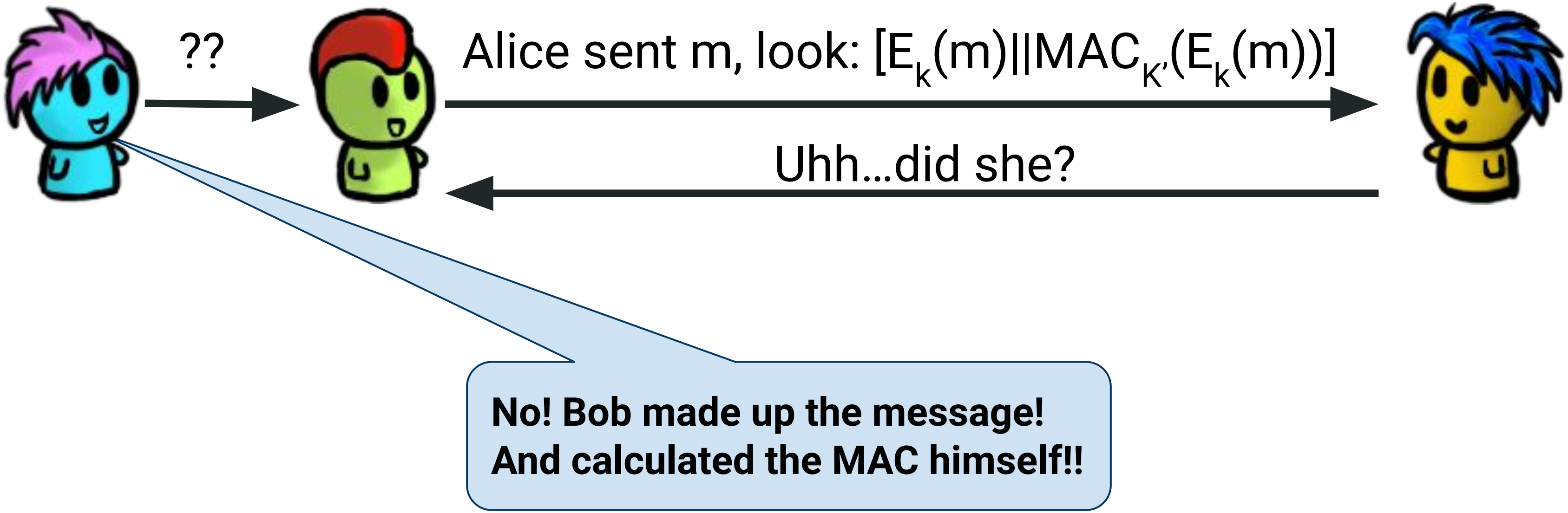
Implications? Repudiation Con’t



Implications? Repudiation Con’t

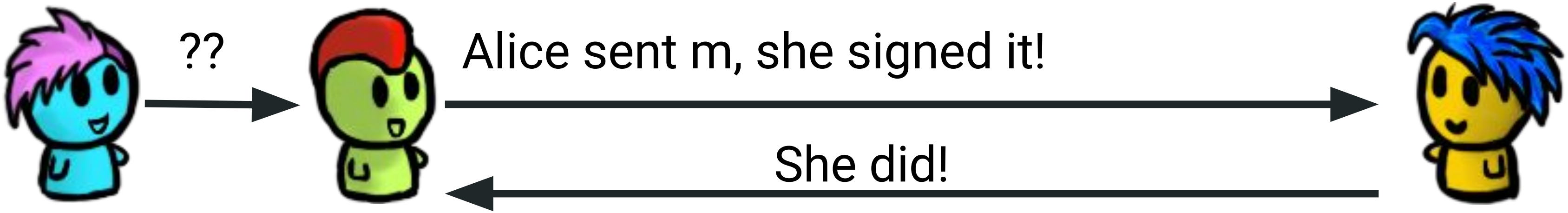


Implications? Repudiation Con't



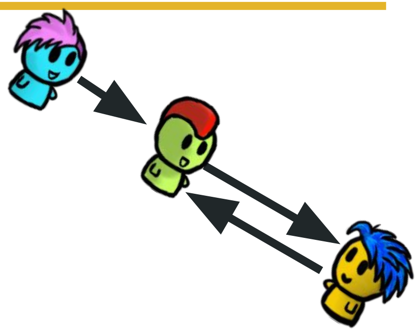
Repudiation Property: For some applications this property is good…others less good (private convos, ecommerce…).

Digital Signatures - For When Repudiation is Bad

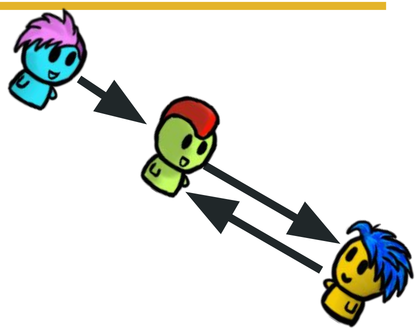


Properties and Goals from Digital Signatures

If Bob receives a message with Alice's digital signature then it should mean:



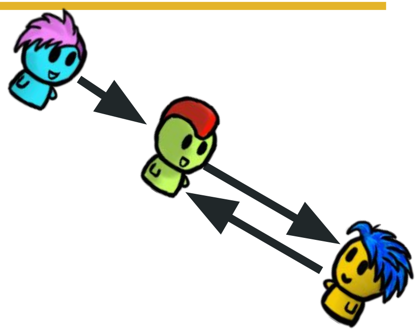
Properties and Goals from Digital Signatures



If Bob receives a message with Alice's digital signature then it should mean:

- Alice sent it (not ), this is like a MAC

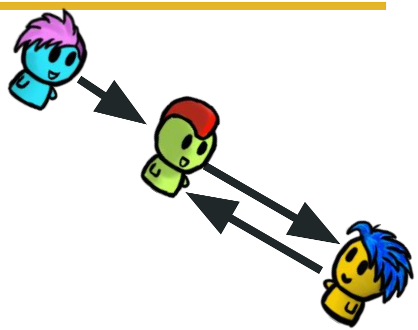
Properties and Goals from Digital Signatures



If Bob receives a message with Alice's digital signature then it should mean:

- Alice sent it (not ), this is like a MAC
- The message has not been altered after sending, MAC

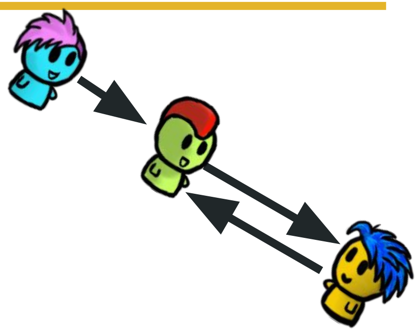
Properties and Goals from Digital Signatures



If Bob receives a message with Alice’s digital signature then it should mean:

- Alice sent it (not ), this is like a MAC
- The message has not been altered after sending, MAC
- The above two properties should be **provable** to a third party, this property is not like a MAC

Properties and Goals from Digital Signatures



If Bob receives a message with Alice's digital signature then it should mean:

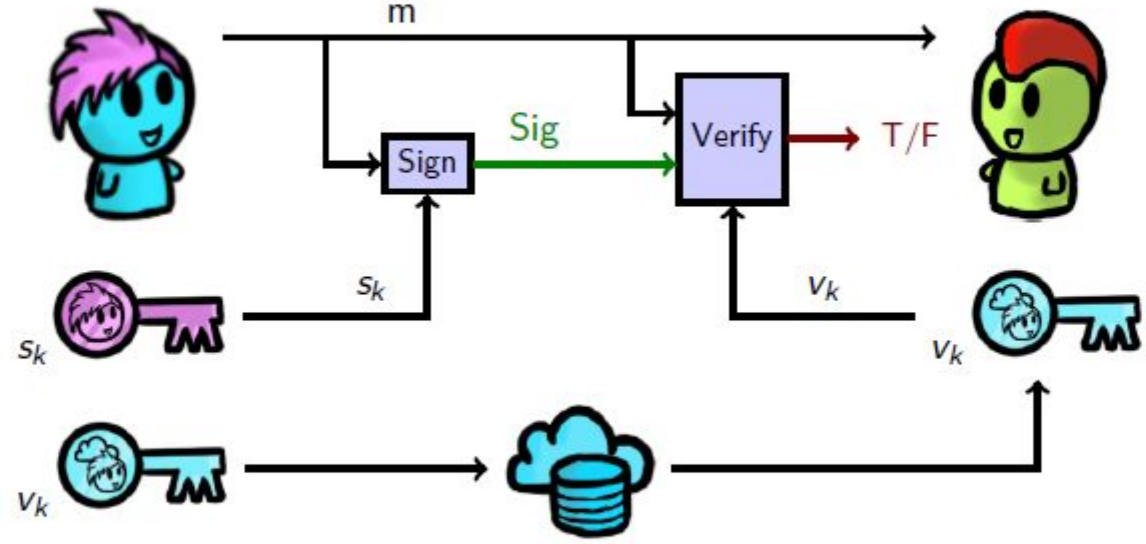
- Alice sent it (not ), this is like a MAC
- The message has not been altered after sending, MAC
- The above two properties should be **provable** to a third party, this property is not like a MAC

Achievable? Use techniques similar to public-key crypto (last class)

Making Digital Signatures

1. Two keys again 
2. Everyone gets the verification key 
3. Alice signs with private signing key 
4. Bob verifies using verification key 
5. If it verifies correctly, success, valid signature

DIgital Signatures at a Glance



Faster Signatures, aka More Hybrids

- Signing large messages, slow
- However, a hash is much smaller than the message…

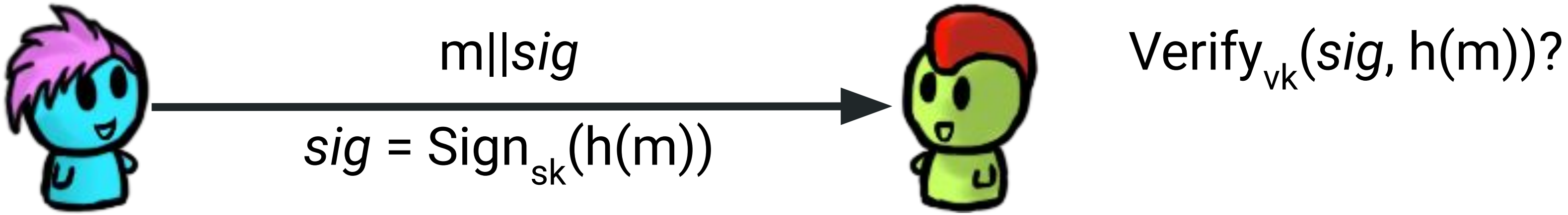
Faster Signatures, aka More Hybrids

- Signing large messages, slow
- However, a hash is much smaller than the message…

$\mathrm{Verify}_{vk}(sig, h(m))$?

Faster Signatures, aka More Hybrids

- Signing large messages, slow
- However, a hash is much smaller than the message…



- Finally, authenticity and confidentiality are separate, you need to include both if you want to achieve both

The Key Management Problem

Q: How can Alice and Bob be sure they're talking to each other?

The Key Management Problem

Q: How can Alice and Bob be sure they're talking to each other?

A: By having each other's verification key!

The Key Management Problem

Q: How can Alice and Bob be sure they're talking to each other?

A: By having each other's verification key!

Q: But how do they get the keys…

The Key Management Problem…Solutions?

Q: But how do they get the keys…

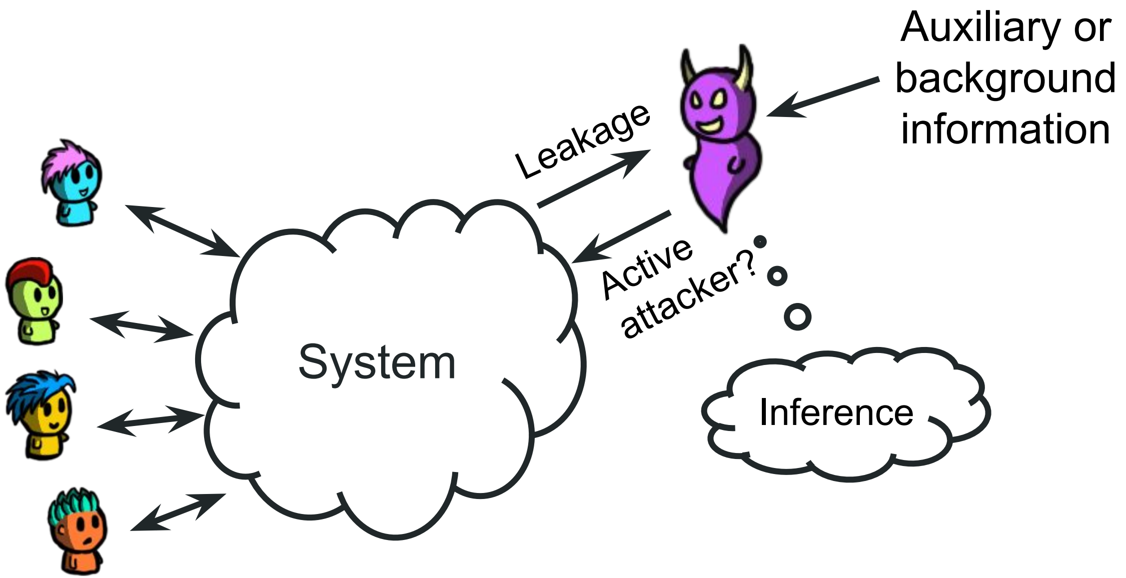
A: Know it personally (manual keying e.g., SSH)

A: Trust a friend (web of trust e.g, PGP)

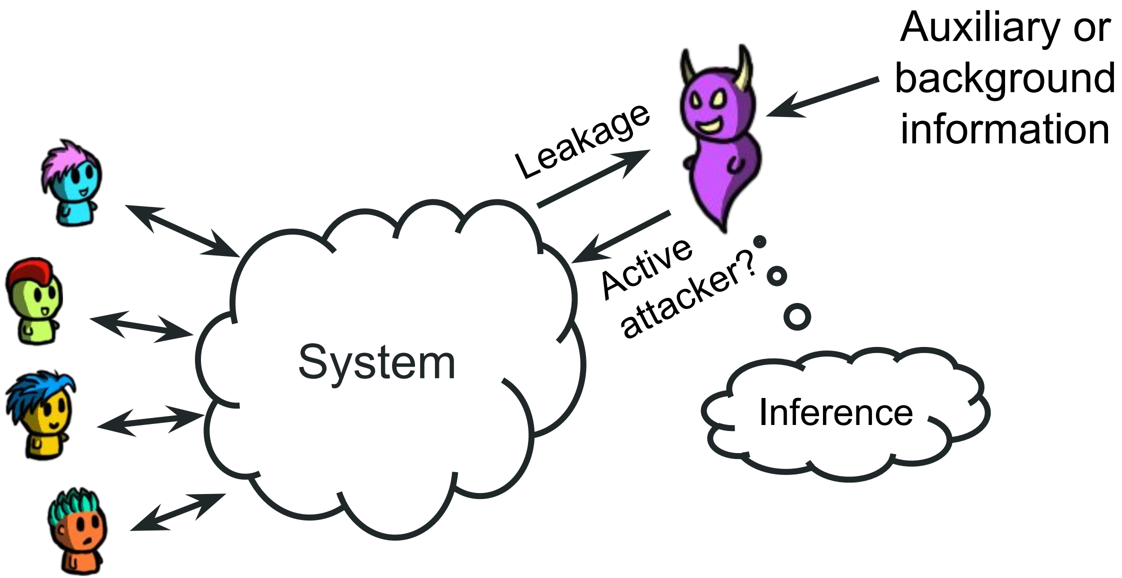
A: Trust some third party to tell them (CAs, e.g., TLS/SSL)

Inference Attacks?

What are inference attacks?

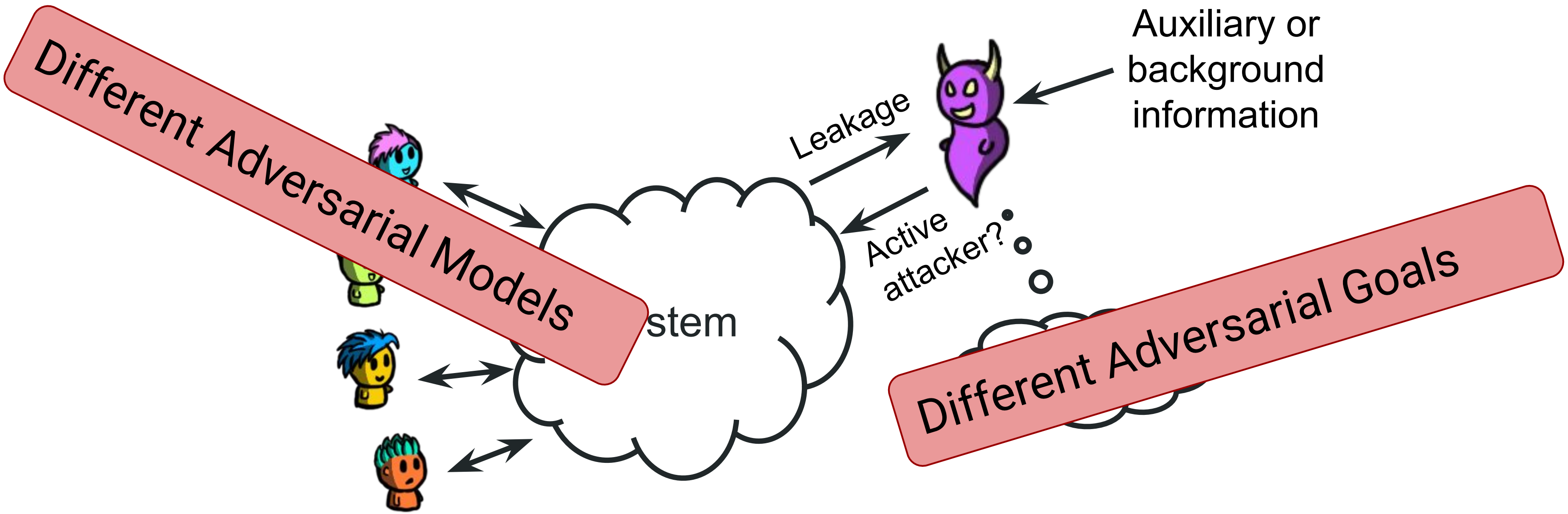


What are inference attacks?



Goal: Learn something (non-trivial) and privacy sensitive from the system

What are inference attacks?



Goal: Learn something (non-trivial) and privacy sensitive from the system

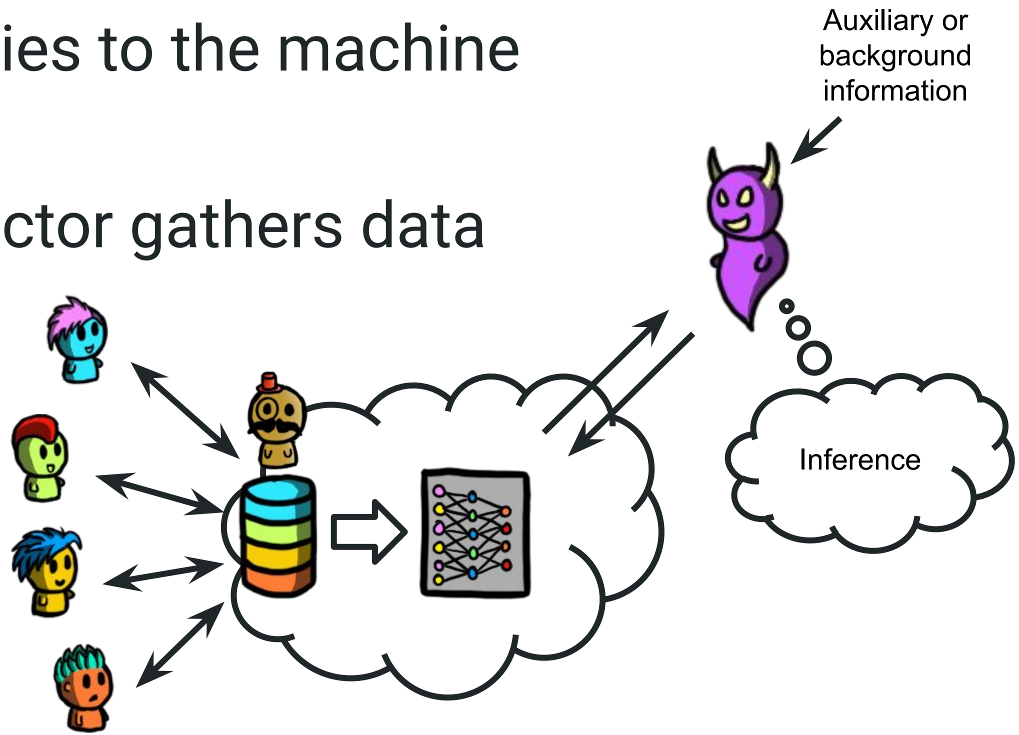
Context for Inference Attacks: The Model

- Attacks generally rely on information “leakage”
- The leakage can be intentional:
  - Sending usage statistics to a service provider (Microsoft, Apple, …)
  - Reporting our location to Google Maps
  - Publishing census data
- Some leakage is unintentional:
  - E.g., side-channels: you saw these earlier!

Attacks can combine all leaked information with auxiliary information to infer non-trivial sensitive data!

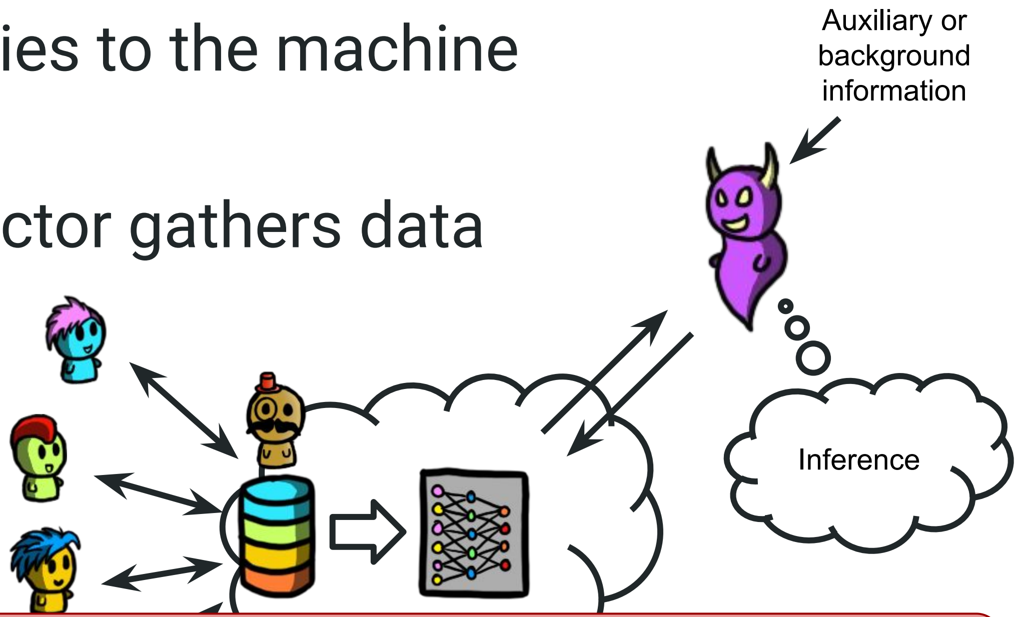
Example: Machine Learning

- **Adversary:** can issue queries to the machine learning model.
- **Functionality:** A data collector gathers data from users and trains a machine learning model with it (they don't intend to leak anything non-trivial to the adversary).



Example: Machine Learning

- **Adversary:** can issue queries to the machine learning model.
- **Functionality:** A data collector gathers data from users and trains a machine learning model with it (they don't intend to leak anything non-trivial

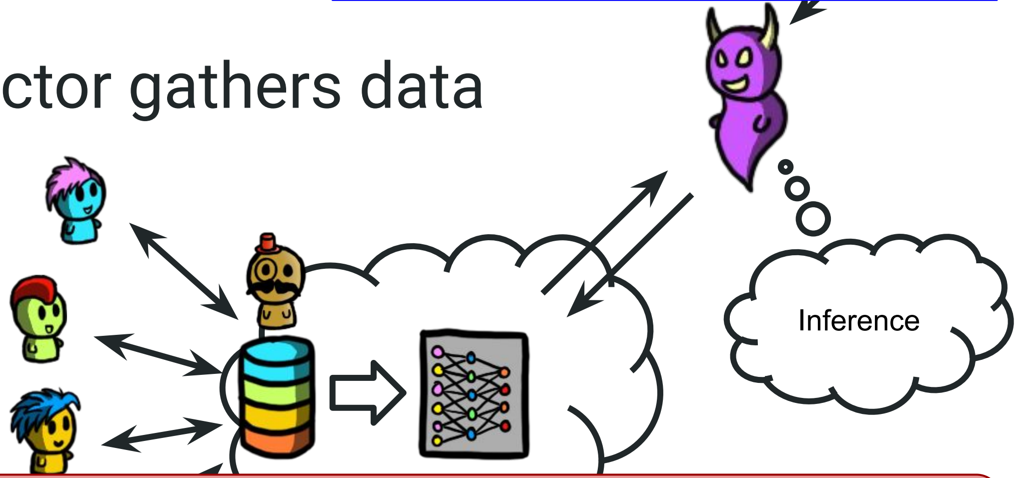


Q: What non-trivial privacy-sensitive information could the adversary infer?

Example: Machine Learning

- **Adversary:** can issue queries to the m
  learning model.
- **Functionality:** A data collector gathers data from users and trains a machine learning model with it (they don't intend to leak anything non-trivial

Leakage:
- Inferences from the ML model

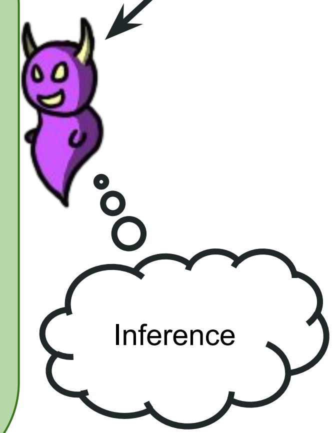


Q: What non-trivial privacy-sensitive information could the adversary infer?

Example 2: Machine Learning

- **Advers**…
  learn…
- **Funct**…
  from…
  mach…
  with i…
  to leak…

Leakage:
- Inferences from the ML model



A:
- Each user’s data (the whole training dataset)
- Whether or not a particular data sample was in the training set
- A general property of the training population
- Given partial data about a user, learn other attributes about the user
- …

Q: What non-trivial privacy-sensitive information could the adversary infer?

Why study inference attacks?

Adversarial Thinking

- Think like an adversary to understand the ***vulnerabilities*** of a system and develop ***protection techniques***.
- When designing inference attacks, we also apply **Kerckhoff's principle** (or Shannon's maxim), adapted to privacy

Adversarial Thinking

- Think like an adversary to understand the ***vulnerabilities*** of a system and develop ***protection techniques***.
- When designing inference attacks, we also apply **Kerckhoff's principle** (or Shannon's maxim), adapted to privacy

Assume the adversary knows how the system works

- there are no hidden parameters other than the users' data
- the adversary can even know some rough distribution that the users' data follows)

Designing a System Aware of Inference Attacks

For any system that relies on users’ data, there are two goals:

- **Utility:** Design a system that provides benefits to its users and the service provider
- **Privacy:** Design a system that provides protection against inference attacks

Q: What are “utility” and “privacy”? How do we “measure” them?

Designing a System Aware of Inference Attacks

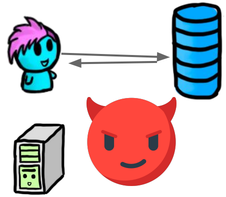
For any system that relies on users' data, there are two goals:

- **Utility:** Design a system that provides benefits to its users and the service provider
- **Privacy:** Design a system that provides protection against inference attacks

Q: What are "utility" and "privacy"? How do we "measu

It's complicated...

Recall, What is privacy?



Technical Privacy

Conceptual Privacy

Legal Privacy

Usable Privacy

What is privacy?

- Useful definition: informational self-determination

"The right of the individual to decide what information about himself should be communicated to others and under what circumstances" (Westin, 1970)

- Privacy is having control over:
  - Who we share our data with
  - Who they can share it with
  - For what purpose they use it
  - Etc.

Quantifying Privacy?

- Protecting the sensitive information e.g., not just data, also meta-data, relationships, timing, whether a user participated in a system, etc.
- Quantifying privacy is very hard

There is **no cure-all metric** for privacy, measuring privacy can be computationally intractable, etc.

Quantifying Privacy: Theoretical Notions

- **Syntactic** notions of privacy: these are computed on the leaked or released data. They are data dependent
  - K-anonymity, l-diversity, t-closeness, etc

Quantifying Privacy: Theoretical Notions

- **Syntactic** notions of privacy: these are computed on the leaked or released data. They are data dependent
  - K-anonymity, l-diversity, t-closeness, etc
- **Semantic** notions of privacy: these are computed on the data release mechanism itself, and they hold regardless of the data (data independent)
  - Mostly Differential Privacy

Quantifying Privacy: Empirical Notions

- The performance of an **inference attack** e.g., the attacker error, accuracy, true positive rate, false positive rate, etc
- Can provide an **upper bound** on privacy

Quantifying Privacy: Empirical Notions

- The performance of an **inference attack** e.g., the attacker error, accuracy, true positive rate, false positive rate, etc
- Can provide an **upper bound** on privacy

Q: Why an upper bound?

Quantifying Privacy: Empirical Notions

- The performance of an **inference attack** e.g., the attacker error, accuracy, true positive rate, false positive rate, etc
- Can provide an **upper bound** on privacy

Q: Why an upper bound?

A: Can't get more privacy if this attack succeeds

Utility and Privacy

Utility

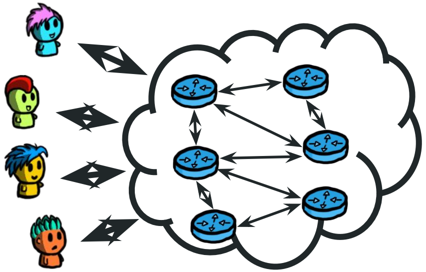
Definition: the benefit that users (and the provider) get from using the system.

Utility

Definition: the benefit that users (and the provider) get from using the system.

Communications system:

- For users: being able to communicate

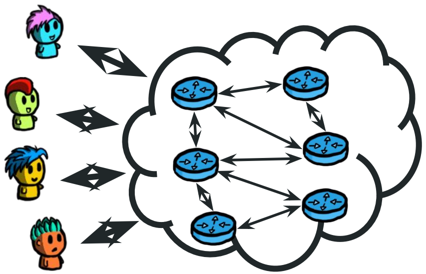


Utility

Definition: the benefit that users (and the provider) get from using the system.

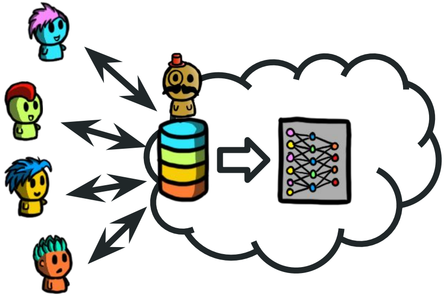
Communications system:

- For users: being able to communicate



Machine learning:

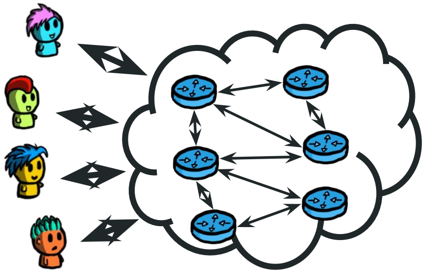
- For participants: maybe they get compensation?
- For data owner: it can sell access to the model for revenue
- Analysts: they pay to get benefits from the model's outputs
- General public: maybe the model outputs are good for society?



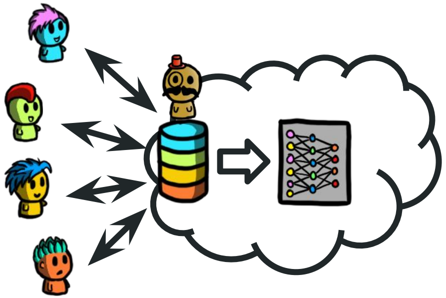
Quantifying Utility

Q: How do we *quantify* utility?

Communications system:



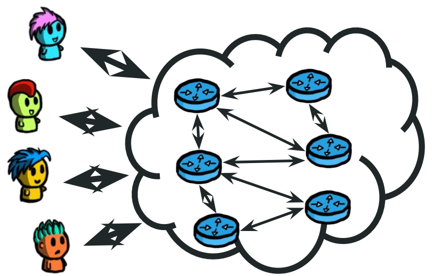
Machine learning:



Quantifying Utility

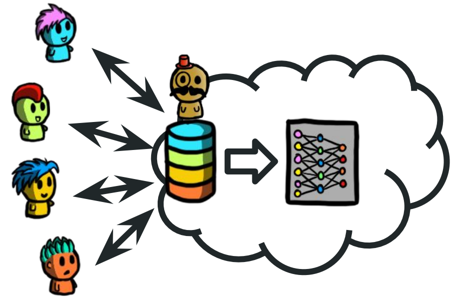
Q: How do we *quantify* utility?

Communications system:



- Low packets dropped
- High bandwidth/throughput
- Low latency/delay…

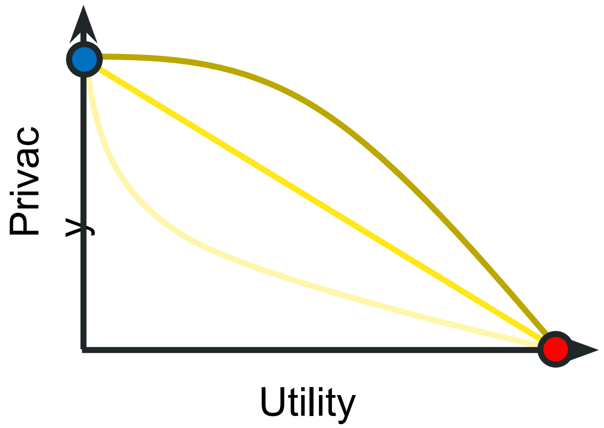
Machine learning:



- Useful model (high test accuracy)
- Unbiased model (low disparity among subpopulations)
- Low computational requirements to build the model
- Fast training algorithm…

The Privacy-Utility trade-off

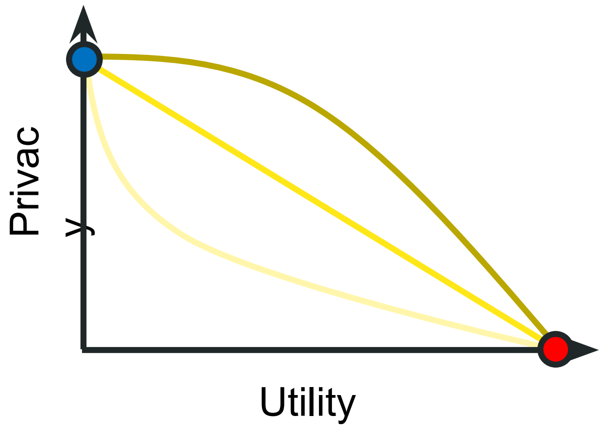
- Given any metric for privacy and for utility, they are usually at odds:



- **Q:** How do you design a system that provides maximum utility?
- **Q:** How do you design a system that provides maximum privacy?
- Designing a system that provides a good privacy-utility trade-off is hard!

The Privacy-Utility trade-off

- Given any metric for privacy and for utility, they are usually at odds:

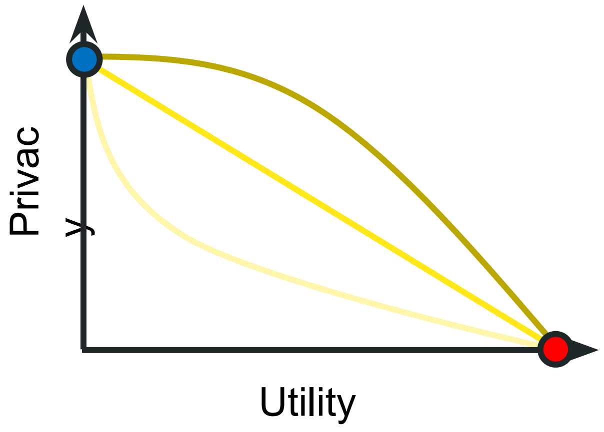


- How do you design a system that provides maximum utility?
  - You design it without privacy in mind
- How do you design a system that provides maximum privacy?

- Designing a system that provides a good privacy-utility trade-off is hard!

The Privacy-Utility trade-off

- Given any metric for privacy and for utility, they are usually at odds:



- How do you design a system that provides maximum utility?
  - You design it without privacy in mind
- How do you design a system that provides maximum privacy?
  - You don't design it
- Designing a system that provides a good privacy-utility trade-off is hard!

Inference Attacks: Goals and Techniques

- As we saw before, the attacker can have different **goals**:
  - Infer data
  - Infer a property of the data
  - Infer the presence (membership) of some data
  - Infer the behavior of a user
  - Infer some attributes of a data sample
  - Infer dependencies among the data
  - …

Inference Attacks: Goals and Techniques

- As we saw before, the attacker can have different **goals**:
  - Infer data
  - Infer a property of the data
  - Infer the presence (membership) of some data
  - Infer the behavior of a user
  - Infer some attributes of a data sample
  - Infer dependencies among the data
  - …
- There are different **techniques** to perform an inference attack:
  - Statistical tools (estimation theory, detection theory, maximum likelihood, Bayesian inference…)
  - Combinatorics
  - Heuristics
  - Machine learning
  - …

Inference Attack Examples

Inference attacks: examples

- Let's see examples of inference attacks with different **goals** and **techniques**.
- You need to understand these attacks, their goal, the leakage they exploit and the techniques they use.

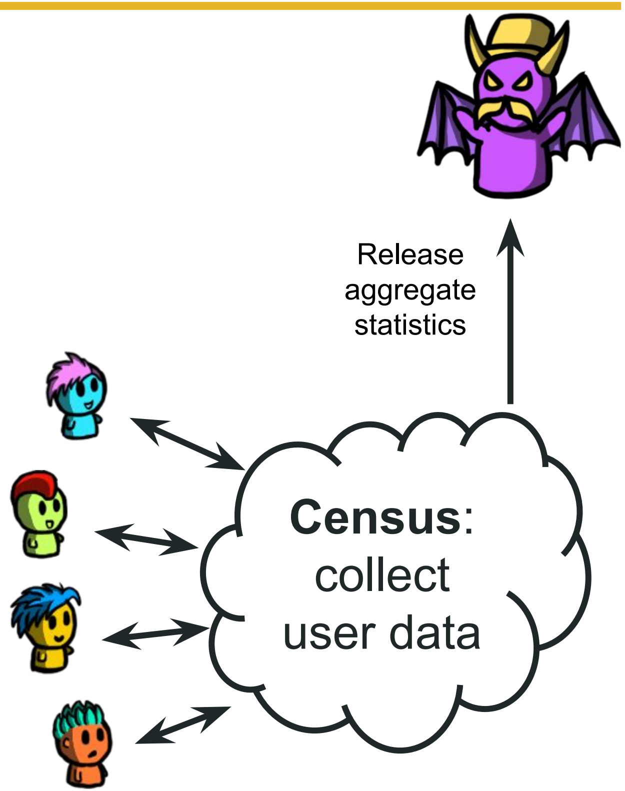
There are:

1. Census reconstruction attacks
2. SQL inference attacks (tracker attacks)
3. Database reconstruction attacks
4. Statistical inference attacks
    - Maximum Likelihood
    - Maximum A-Posteriori
5. De-anonymization attacks
6. Side-channel attacks
7. ML Inference attacks
8. Linking attacks

Census Reconstruction Attacks

Census Reconstruction Attacks

- A census involves collecting lots of privacy-sensitive data.
- Some useful aggregate statistics are released.
- The adversary tries to infer (reconstruct) some individuals' data.

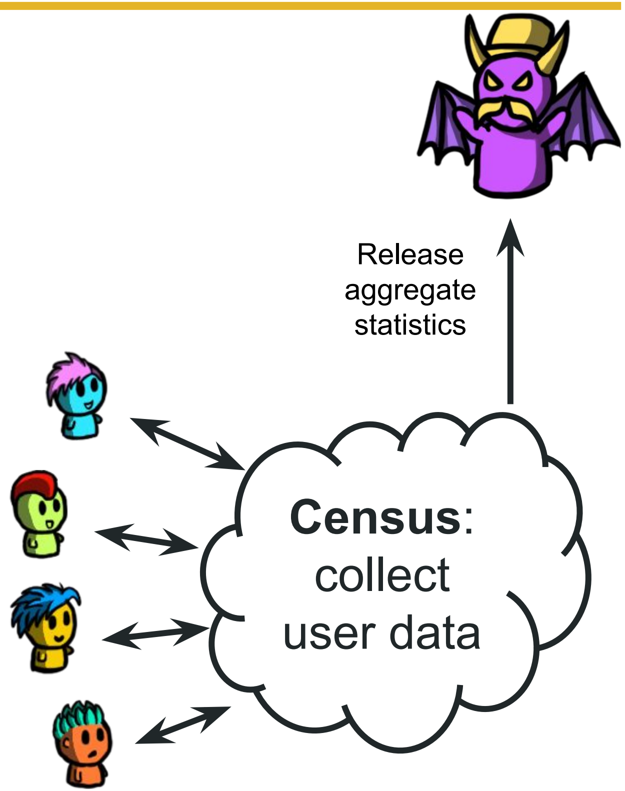


Census Reconstruction Attacks

- A census involves collecting lots of privacy-sensitive data.
- Some useful aggregate statistics are released.
- The adversary tries to infer (reconstruct) some individuals' data.
- Example:

Background data: adversary knows a participant that self-identifies as white is 35 years old.
Released aggregates:

	COUNT	AGE MEAN
Total population	4	24
White	2	26
Asian	2	22



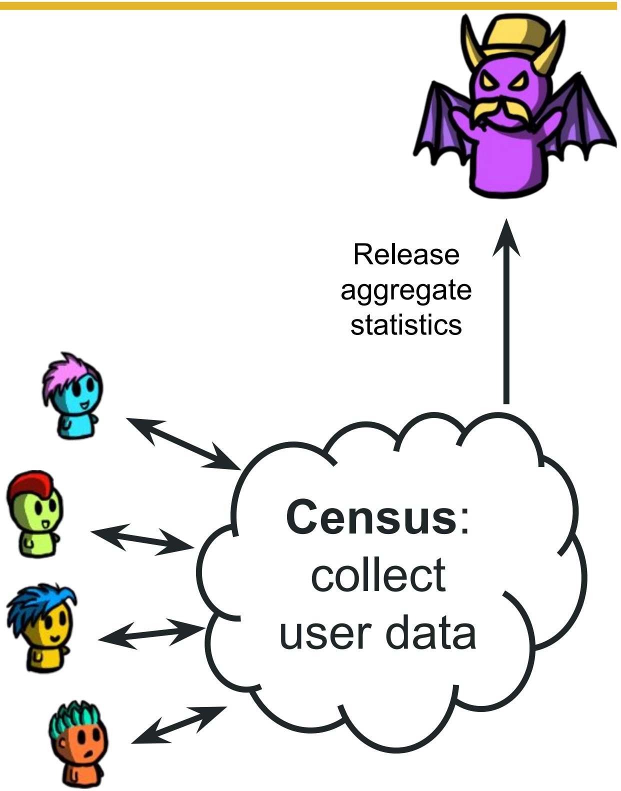
Census Reconstruction Attacks

- A census involves collecting lots of privacy-sensitive data.
- Some useful aggregate statistics are released.
- The adversary tries to infer (reconstruct) some individuals' data.
- Example:

Background data: adversary knows a participant that self-identifies as white is 35 years old.
Released aggregates:

	COUNT	AGE MEAN
Total population	4	24
White	2	26
Asian	2	22

Q: Can you guess the age and self-identified race of every participant?



Census Reconstruction Attacks

- A census involves collecting lots of privacy-sensitive data.
- Some useful aggregate statistics are released.
- The adversary tries to infer (reconstruct) some individuals' data.
- Example:

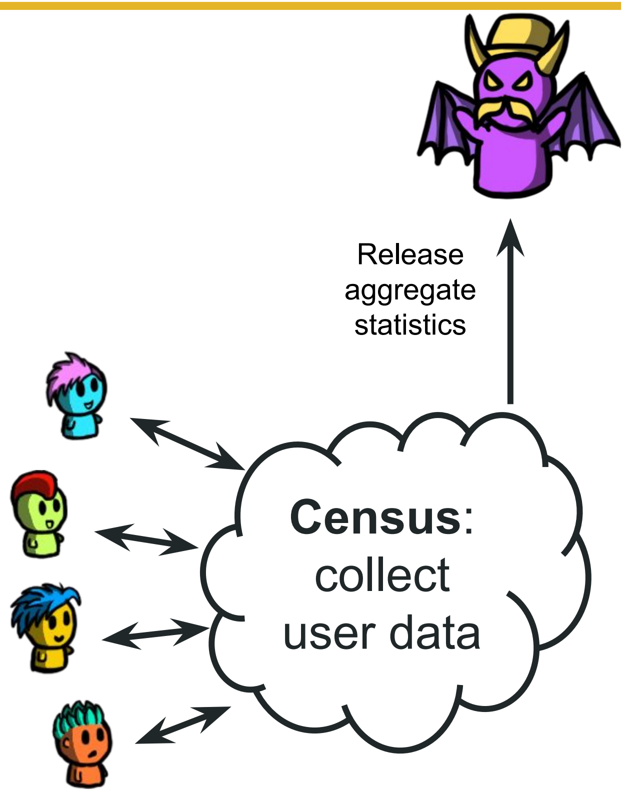
Background data: adversary knows a participant that self-identifies as white is 35 years old.

Released aggregates:

	COUNT	AGE MEAN
Total population	4	24
White	2	26
Asian	2	22

Q: Can you guess the age and self-identified race of every participant?

A: W1=17, W2=35, A1=21, A2=23



Census reconstruction attacks

- Another example, no background information:

Q: Can you guess the self-identified race, age, and marital status?

	COUNT	AGE MEAN	AGE MEDIAN
Total population	4	37.5	35.5
White	2	42.5	42.5
Asian	2	32.5	32.5
Single	1	25	25
Married	3	41.66	31

Census reconstruction attacks

- Another example, no background information:

Q: Can you guess the self-identified race, age, and marital status?

	COUNT	AGE MEAN	AGE MEDIAN
Total population	4	37.5	35.5
White	2	42.5	42.5
Asian	2	32.5	32.5
Single	1	25	25
Married	3	41.66	31

A: If you **assume the single person is Asian**, A1=25, then A2=40.
One white has to be W=31 (because that's the median of married), and the other white is W=54.
These values meet the total population age median.

Census reconstruction attacks

- Another example, no background information:

Q: Can you guess the self-identified race, age, and marital status?

	COUNT	AGE MEAN	AGE MEDIAN
Total population	4	37.5	35.5
White	2	42.5	42.5
Asian	2	32.5	32.5
Single	1	25	25
Married	3	41.66	31

A: If you assume the single person is Asian, A1=25, then A2=40.
One white has to be W=31 (because that's the median of married), and the other white is W=54.
These values meet the total population age median.

If you **do the same assuming the single is White**, you get W1=25, W2=54, A1=31, A2=34, which does not meet the age median result, so it can't be true.

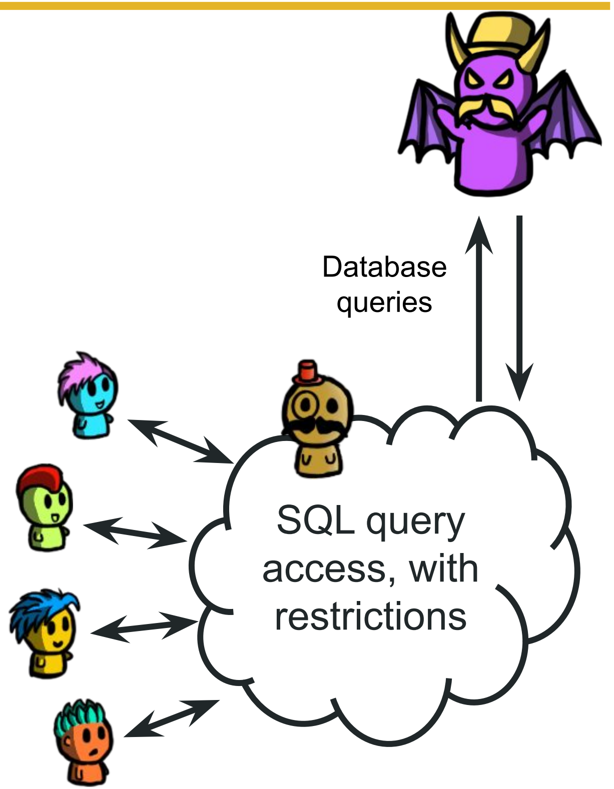
SQL Query Attacks

SQL query attacks

- A data collector creates a relational database (table) with data from different clients.
- An adversary can issue SQL queries to gather data from the table.
- The database management system allows queries with the following syntax:

SELECT SUM(ATTRIBUTE) FROM (TABLE) WHERE (CONDITION)

- However, any queries that match less than X entries or more than N-X entries are discarded.



SQL query attacks: example

Name	Age	Position	Salary
Alice	40	full time	120k
…	…	…	…
Carol	…	…	…
…	…	…	…

- The table Employees has four attributes:
  - Names are unique
  - Ages are between 18 and 65
  - Position is either ‘full time’ or ‘part time’
  - Salaries are between 50k and 500k
- You know Carol is in the dataset, and that around 50% of the people in the dataset are ‘full time’.
- There are N records in the dataset; any query that matches less than $\frac{N}{10}$ or more than $\frac{9N}{10}$ entries *is discarded.*
- Can you recover Carol’s salary? How many queries do you need?

SELECT SUM(ATTRIBUTE) FROM (TABLE) WHERE (CONDITION)

SQL query attacks: solution

- There are N records in the dataset; any query that matches less than $\frac{N}{10}$ or more than $\frac{9N}{10}$ entries *is discarded.*

Name	Age	Position	Salary
Alice	40	full time	120k
…	…	…	…
Carol	…	…	…
…	…	…	…

Solution:
Q1=SELECT SUM(Salary) FROM Employees WHERE (Position='full time' OR Name=Carol)
Q2=SELECT SUM(Salary) FROM Employees WHERE (Position='full time' AND Name!=Carol)
Salary=Q1-Q2

If Carol is part time:

		Q1	Q2	Q1-Q2
Full time				
Part time				
	Carol			

If Carol is full time:

		Q1	Q2	Q1-Q2
Full time				
	Carol			
Part time				

Q1-Q2 always gets Carol's salary!

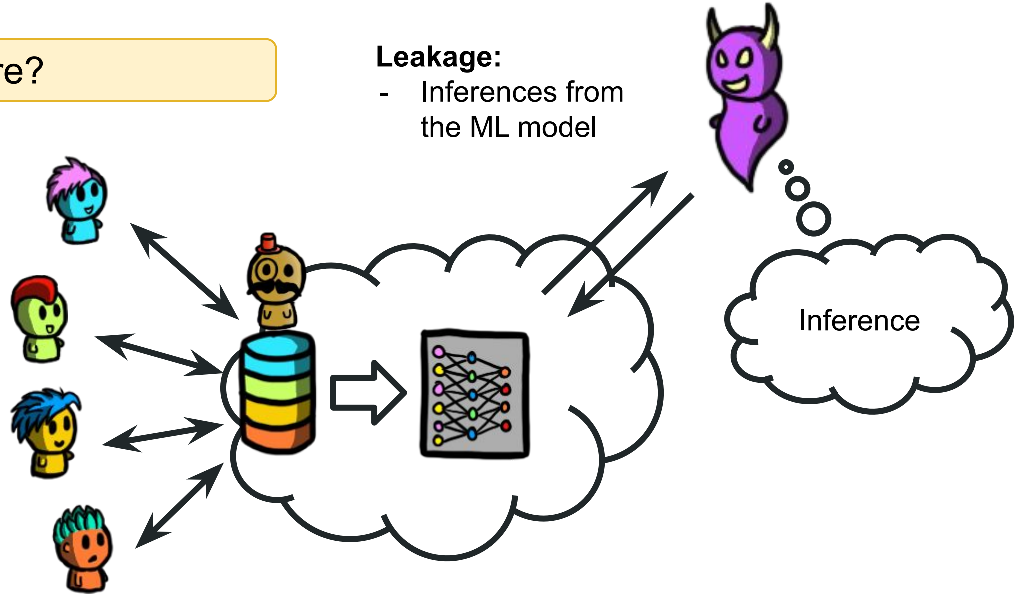
SQL query attacks:

- The lesson is: even if the result of a query is harmless (too general), the combination of two or more queries can be very dangerous (very specific).
- Placing restrictions on individual queries, while still reporting exactly values, does not work.
- When coming up with SQL query attacks in this setting:
  - Look for an attribute that you can use to make sure you always bypass the restriction so that the query goes through.
  - After you design the queries, check that they get the desired value regardless of the values of other attributes in the dataset (e.g,. whether Carol was full or part time in the previous example)

Inference attacks in Machine Learning

- There are many possible inference attacks in ML.
- Think about the adversary **goals** and possible **techniques**

Q: What could be an inference here?



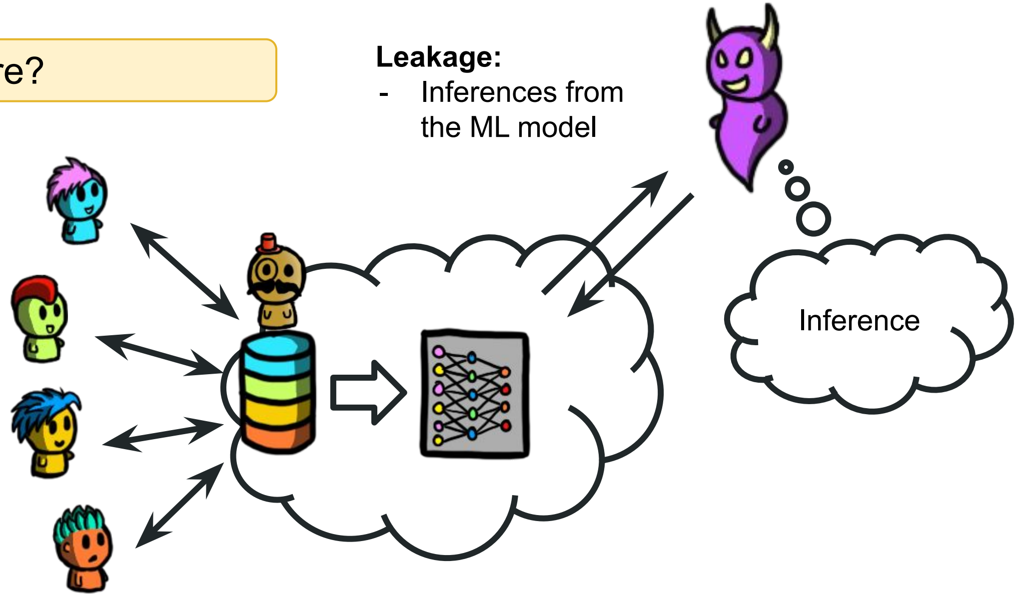
ML Attacks

Inference attacks in Machine Learning

- There are many possible inference attacks in ML.
- Think about the adversary **goals** and possible **techniques**

Q: What could be an inference here?

A:
- Membership inference
- Attribute inference (parts of a data sample)
- Property inference (property of the whole training set)
- Reconstruction attack (infer a whole training set)
- …



Cryptography done…

…for now.

Presenter Instructions

- Send me slides before class (30 minutes before)
- You can use figures etc. from the paper (with attribution)
- Make your own slides
  - Yes, even if it is a USENIX paper and you can download the authors slides
- Practice your timing
- Prepare some discussion prompts

Presentations Proto-Rubric

- Slides quality (appropriate use of space, lack of typos, etc)
- Speaking (audible, pacing/speed, use of space/not distracting)
- Organization (good structure, all the important parts, impact)
- Presenter (variation in voice, eye contact, movement, humour, other?)
- **Timing**
- Discussion facilitation (prepared with sufficient background information to achieve)

Paper Review Proto-Rubric

- (0,1,2) Included all required attributes
- (0,1,2) Accurate [for each of the required attributes]
- (0,1,2) Insightful [for each of the required attributes]

Presenter Feedback Proto-Rubric

- (0,1,2) Included all required attributes
- (0,1,2) Accurate [for each of the required attributes]
- (0,1,2) Helpful [for each of the required attributes]

Required attributes:

- Timing
- Slides
- Presentation (speaking, engagement, etc)