BlindBox: Deep Packet Inspection Over Encrypted Traffic

Justine Sherry, Chang Lan, Raluca Ada Popa, Sylvia Ratnasamy
UC Berkeley

(Work under submission).

Deep Packet Inspection (DPI)

In-network devices which inspect and modify packet *payloads* to enforce security policies.

Intrusion Prevention

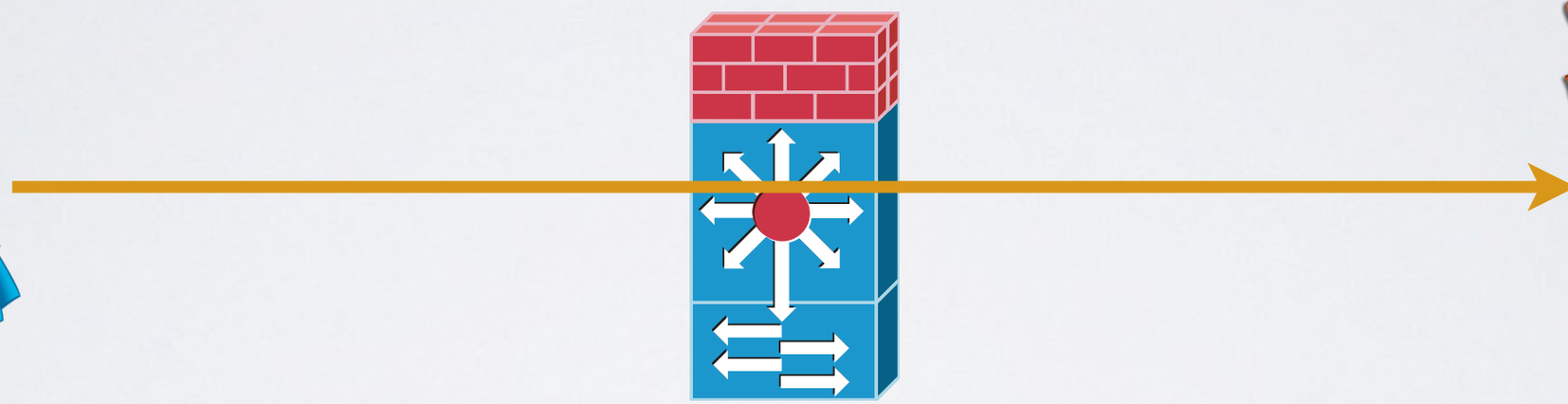
Parental Filtering

Exfiltration Detection

Increasingly offered as "network services." (e.g. NFV, APLOMB)

Alice and Bob

Alice



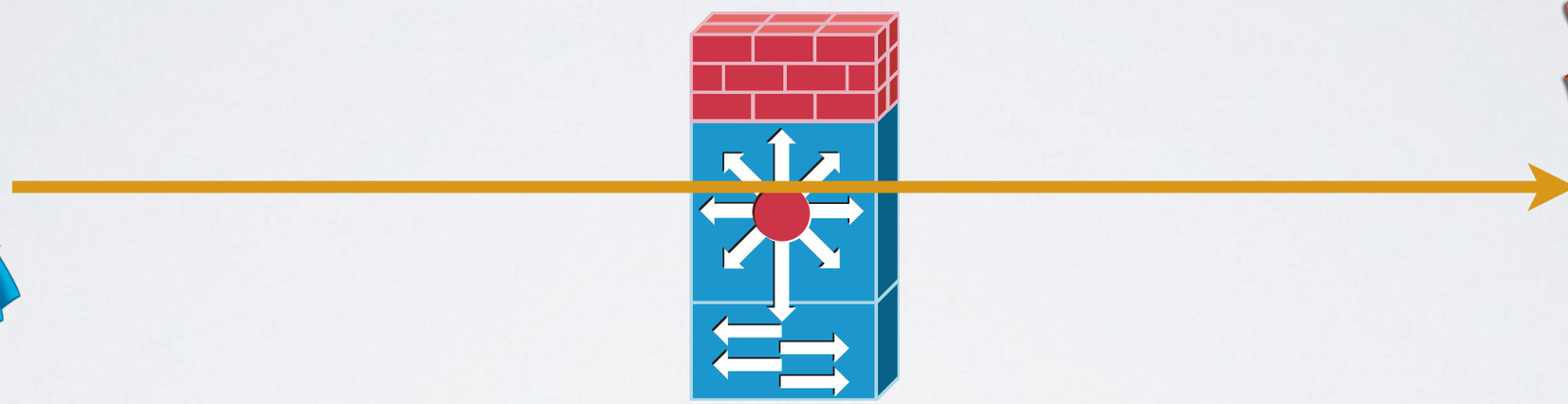
Bob

BLACKLIST
WAREZ
HACKS
%

CONNECTIONS
Alice:Bob ALLOW

McAfee

DPI Usage Today: HTTP



Alice

Bob

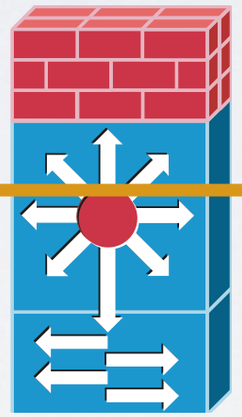
BLACKLIST
WAREZ
HACKS
%

CONNECTIONS

DPI Usage Today: HTTP

To: Bob From:Alice
Hello!

Alice



Bob

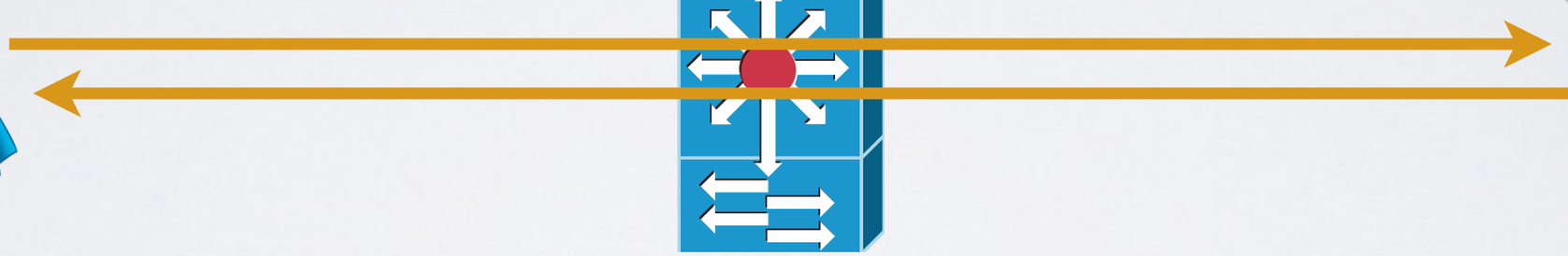
BLACKLIST
WAREZ
HACKS
%

CONNECTIONS
Alice:Bob ALLOW

McAfee

DPI Usage Today: HTTP

To: Alice From:Bob
Hello!



Alice

Bob

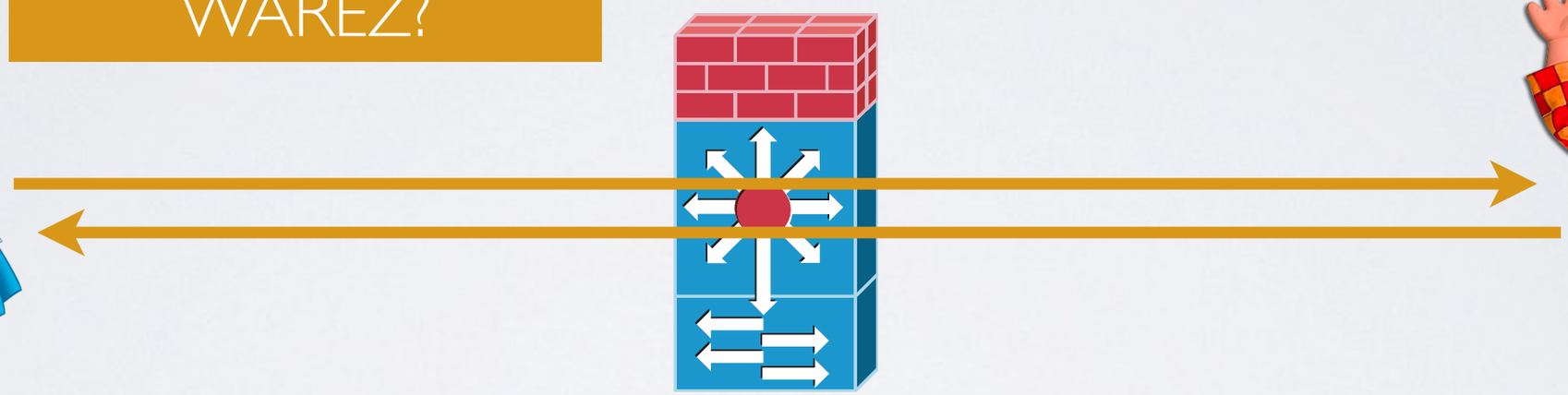
BLACKLIST
WAREZ
HACKS
%

CONNECTIONS
Alice:Bob ALLOW

McAfee

DPI Usage Today: HTTP

To: Bob From:Alice
Want some
WAREZ?



Alice

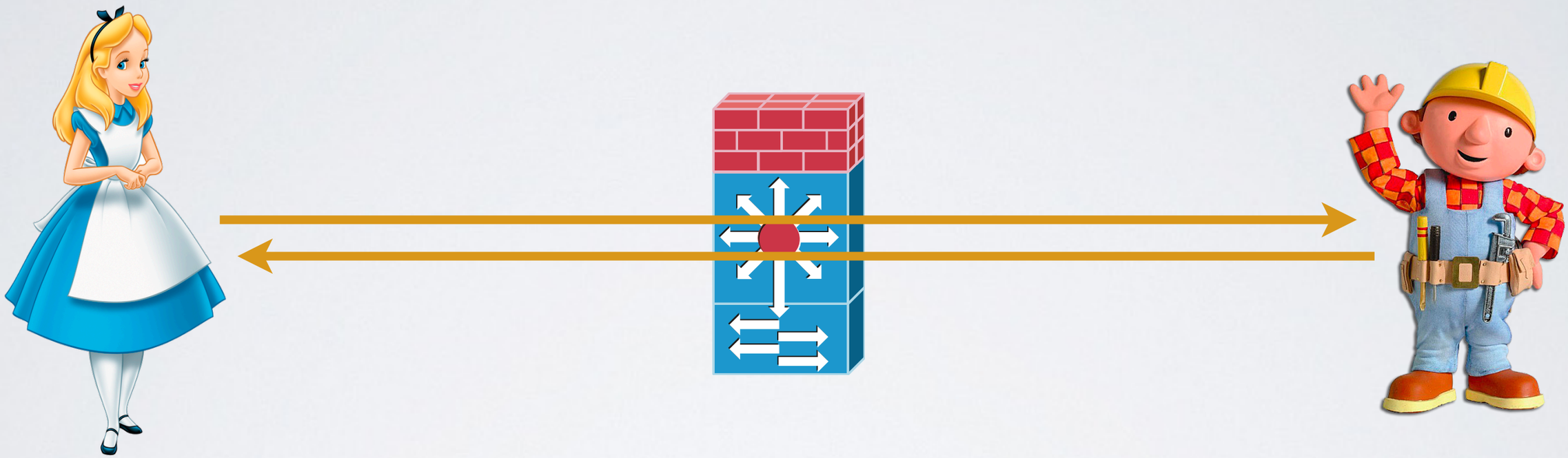
Bob

BLACKLIST
WAREZ
HACKS
%

CONNECTIONS
Alice:Bob ALLOW

McAfee

DPI Usage Today: HTTP



Alice

Bob

BLACKLIST
WAREZ
HACKS
%

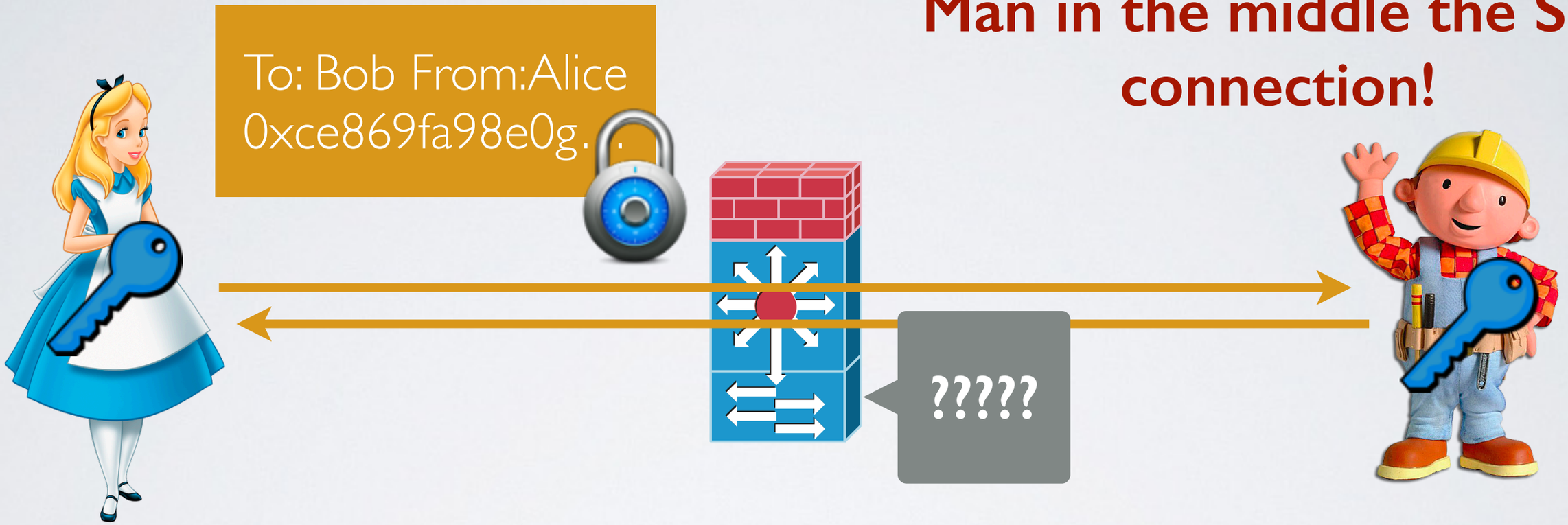
CONNECTIONS
Alice:Bob DENY

McAfee

Many users are switching to HTTPS, specifically to protect their privacy against eavesdroppers.

DPI Usage Today: HTTPS

State of the art solution: Man in the middle the SSL connection!



BLACKLIST
WAREZ
ATTACK
MAD HATTER

CONNECTIONS
Alice:Bob ALLOW

McAfee

BlindBox: Goal

- Alice and Bob have two very conflicting requirements!
  - **Privacy.**
  - **In-network functionality.**

- Can they have their cake and eat it too?

Short answer: *yes!*

BlindBox Functionality

- The first system to allow DPI middle boxes like IPS and Parental Filtering to operate over traffic without granting the ability to decrypt the entire payload.

- "Principle of least privilege": the middle box learns only what it needs to know to detect an attack or match.

Can't functional encryption solve this?

- Existing schemes don't fit our needs:
  - Wrong security model: all parties learn all of the middlebox rules
  - Missing functionality: no approach to address rules which are regular expressions
  - Prohibitive performance: Performing IDS detection over a single packet requires over 1 day of computation on our servers!*

*J. Katz, A. Sahai, B. Waters. "Predicate Encryption Supporting Disjunctions, Polynomial Equations, and Inner Products." EUROCRYPT 2008.

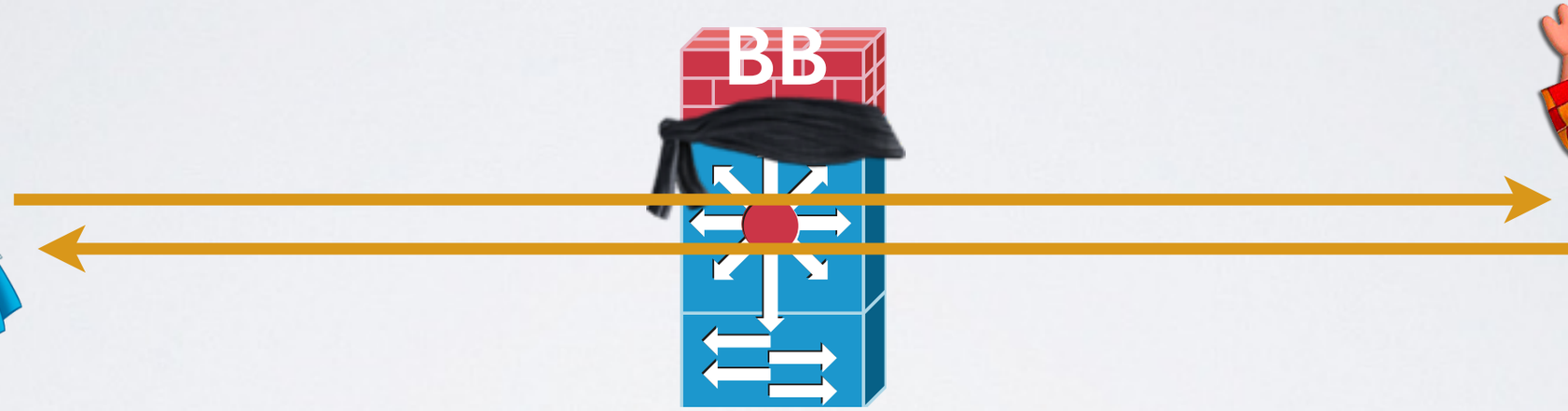
Threat Model Summary: Actors and Constraints

McAfee

- Alice and Bob (Clients):
  - Users who want to protect their privacy from the MB. Also want protection from each other, ie, that their traffic be scanned by the middlebox.
  - Requirement: at least one client must be honest.
- Middlebox (MB):
  - "Honest but curious" network operator who provides an inspection service.
- McAffee ("Rule Generator"):
  - Trusted by MB and Clients to generate rules.
  - Does not have the power to actually observe/manipulate client traffic.

Strawman Approach

Has many security holes, but gets one thing right: searchable encryption.



Alice

Bob

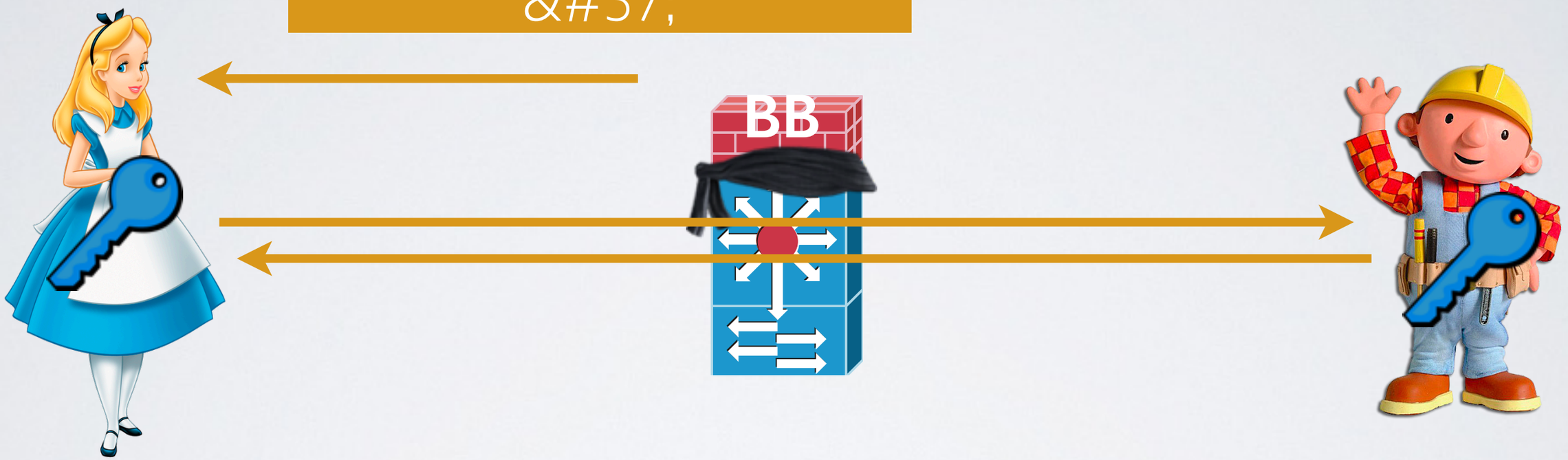
BLACKLIST
WAREZ
HACKS
%

CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

Rules: WAREZ, HACKS, %



Alice

Bob

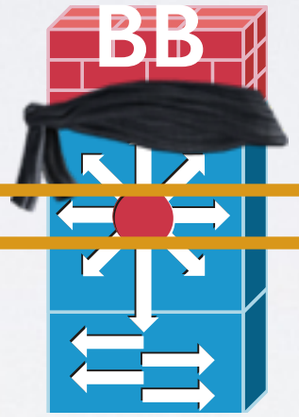
BLACKLIST
WAREZ
HACKS
%

CONNECTIONS
Alice:Bob ALLOW

Strawman Approach

Rules: 0xeaf345, 0x43aa, 0x678ea3

Deterministic AES; no IV, no Salt



Alice

Bob

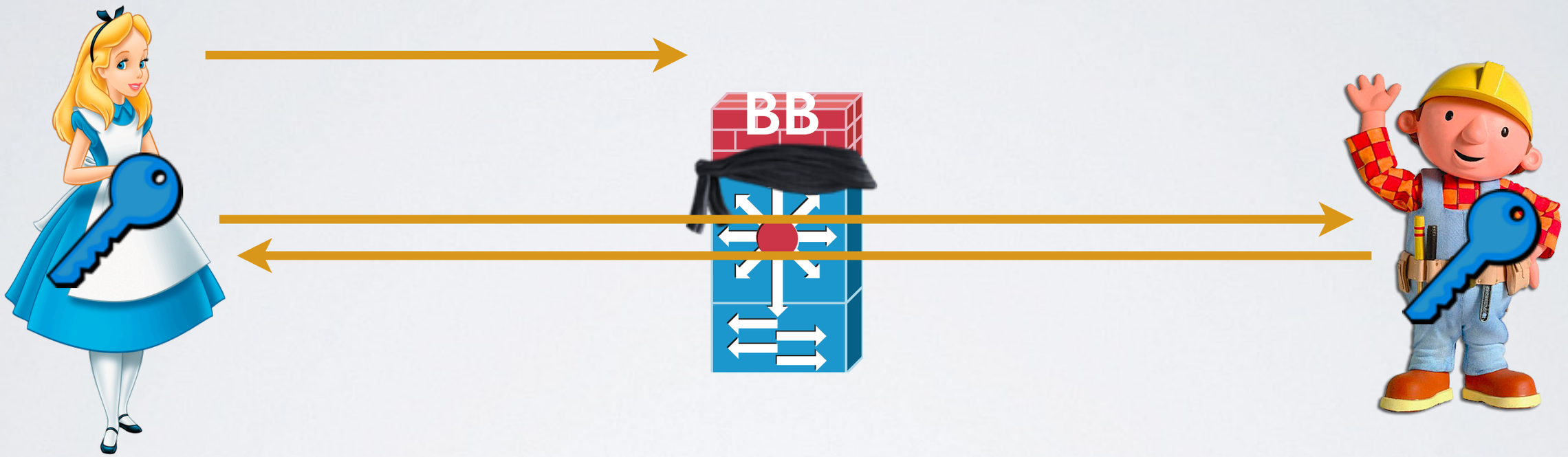
BLACKLIST
WAREZ
HACKS
%

CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

Rules: 0xeaf345, 0x43aa, 0x678ea3



Alice

Bob

BLACKLIST
WAREZ: 0xeaf345…
HACKS: 0x43aa…
%: 0x678ea3…

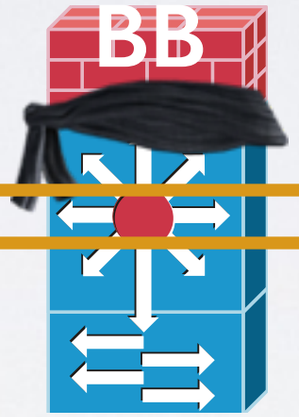
CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

To: Bob From:Alice
Would you like some CAKE?

0xe90326



Alice

Bob

BLACKLIST
WAREZ: 0xeaf345..
HACKS: 0x43aa..
%: 0x678ea3...

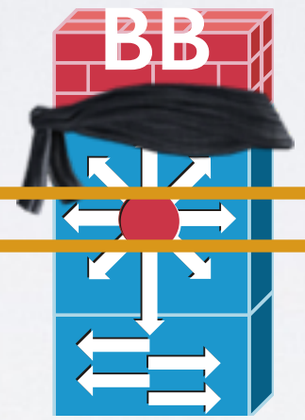
CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

To: Bob From:Alice
Would you like some CAKE?

0x592aa5



Alice

Bob

BLACKLIST
WAREZ: 0xeaf345..
HACKS: 0x43aa..
%: 0x678ea3...

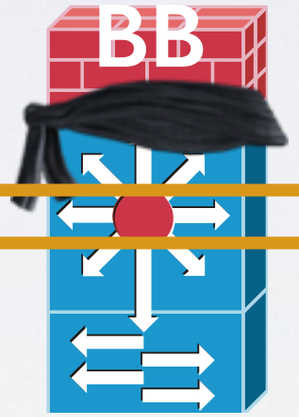
CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

To: Bob From:Alice
Would you like some CAKE?

…etc



Alice

Bob

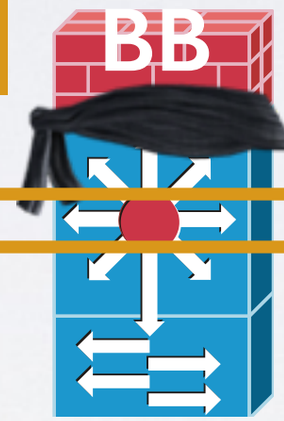
BLACKLIST
WAREZ: 0xeaf345…
HACKS: 0x43aa…
%: 0x678ea3…

CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

To: Bob From:Alice
Would you like some CAKE?



Alice

Bob

BLACKLIST
WAREZ: 0xeaf345..
HACKS: 0x43aa..
%: 0x678ea3...

CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

Alice

Bob

BLACKLIST
WAREZ: 0xeaf345.
HACKS: 0x43aa..
%: 0x678ea3…

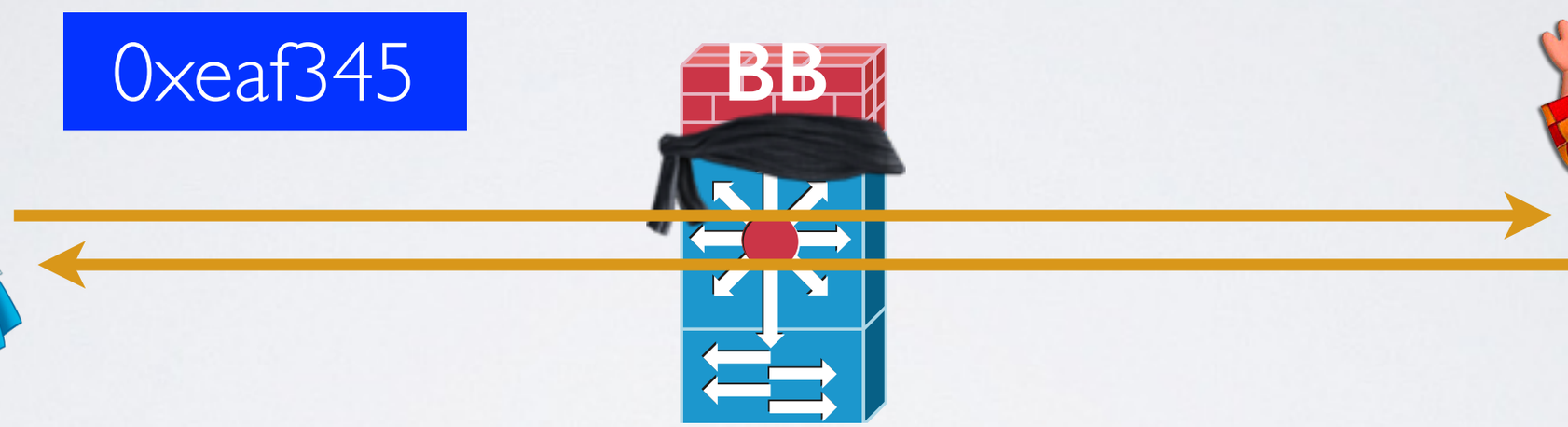
CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

To: Bob From:Alice
Would you like
some WAREZ?

0xeaf345



Alice

Bob

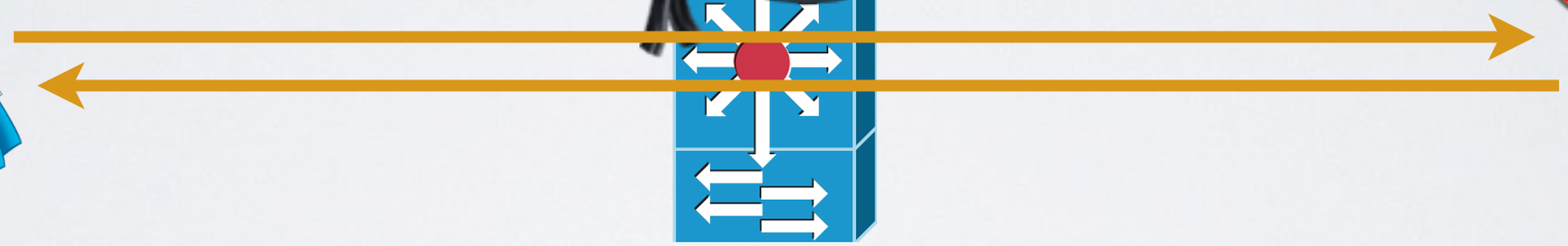
BLACKLIST
WAREZ: 0xeaf345…
HACKS: 0x43aa…
%: 0x678ea3…

CONNECTIONS
Alice:Bob ALLOW

McAfee

Strawman Approach

To: Bob From:Alice
Would you like
some WAREZ?



Alice

Bob

BLACKLIST
WAREZ: 0xeaf345.
HACKS: 0x43aa..
%: 0x678ea3...

CONNECTIONS
Alice:Bob DENY

McAfee

How many bugs did you spot in our Strawman?

Let's fix it.

What was good about the Strawman?

The IDS only learns the decrypted value of the text iff there exists a *rule* for that text.

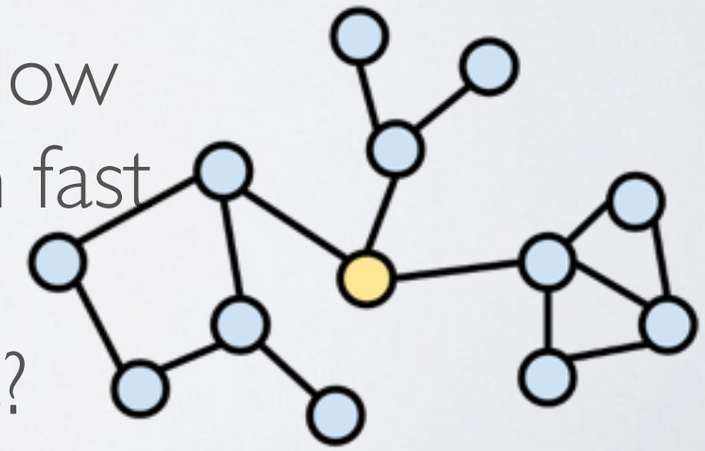
Hence, only text which is "suspicious" can be read by the IDS. The rest remains encrypted.

Fixing Bug #1

- What if there are duplicate substrings in the flow? Won't deterministic encryption leak that there are multiple matches, even for substrings that aren't in the ruleset?

Solution: Just add Salt!

Challenge: How to do so with fast MB data structures?



Fixing Bug #2

- If Alice knows what all the rules are, doesn't she know how to evade detection now?*
  - Also, many IDS rules are *trade secrets* that they are unwilling to share with users/vendors.
- Solution: Yao's Garbled Circuits + Oblivious Transfer

*V. Paxson. "Bro: A System for Detecting Network Intruders in Real Time." Computer Networks 1999.

Fixing Bug #2

- Result:
  - Middlebox learns the encrypted value of the rules, without learning Alice's key.
  - Alice doesn't learn what the rules are.
  - Operation only works if Middlebox's rules have been signed by the rule generator.

Fixing Bug #3

- Some rules are regular expressions (or even in some cases, scripts), not exact matches.

- Solution: "Probable Cause Encryption", a new form of attribute based encryption (ABE).

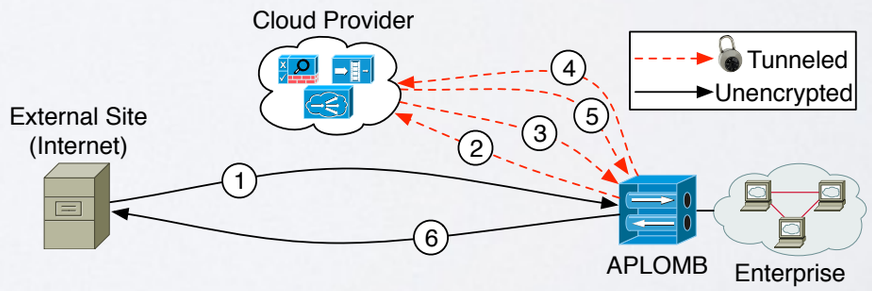
- Key Idea: A second protocol by which MB gains the ability to decrypt the payload only if a set of exact matches have already been detected.

More details in our paper!

- Optimizations to reduce bandwidth overhead.
- Details on GC + OH Transfer.
- How to do fast matching at the middlebox, despite random salts.
- Rule generation, regular expressions, probable cause decryption…

Evaluation Highlights

- Three main performance figures:
  - **Detection Time:** competitive with existing IDSes
    - 186Mbps with BB (Snort Achieves 85Mbps)
  - **Transmission Time:** practical overhead
    - Page load completion time increases by 0.15-1x
  - **Setup Time:** not yet competitive
    - 414s for 3000 rules.



Fine for NFV & APLOMB where connections are persistent.

Conclusion

- BlindBox is the first system to allow network appliances to perform *deep packet inspection* over traffic without needing to decrypt the entire stream.
- Alice and Bob can "have their cake and eat it too", keeping the communications **private**, while receiving the **benefits of network services** like IDS.

Old/Backup Slides

BlindBox Wishlist & Future Work

- Faster setup time (< 1 second) setup time.
- "All or nothing property": leaks only whether or not a *complete* rule matched (not substrings)
  - "Maliciously" -> "Maliciou" + "iciously"
- General regular expression support.

Generalizing to More Middleboxes

- Follow-on work looks at cloud case in general and more middle boxes — including firewalls, NATs, proxies, etc.

- C. Lan, J. Sherry, R. A. Popa, S. Ratnasamy. "Securely Outsourcing Middleboxes to the Cloud."

Non-Usage Scenario

Charlie

- Charlie is a political dissident in a country which deploys DPI devices for *censorship*.
- Charlie is afraid of political repercussions for the things that he reads and writes on the web.
- Charlie should *not* opt-in to BlindBox.
- Even if he trusts his Rule Generator, there is no guarantee that the Rule Generator has not been co-opted by the government!
- Charlie should use a strong encryption scheme instead.

Usage Scenario # 1

Alice

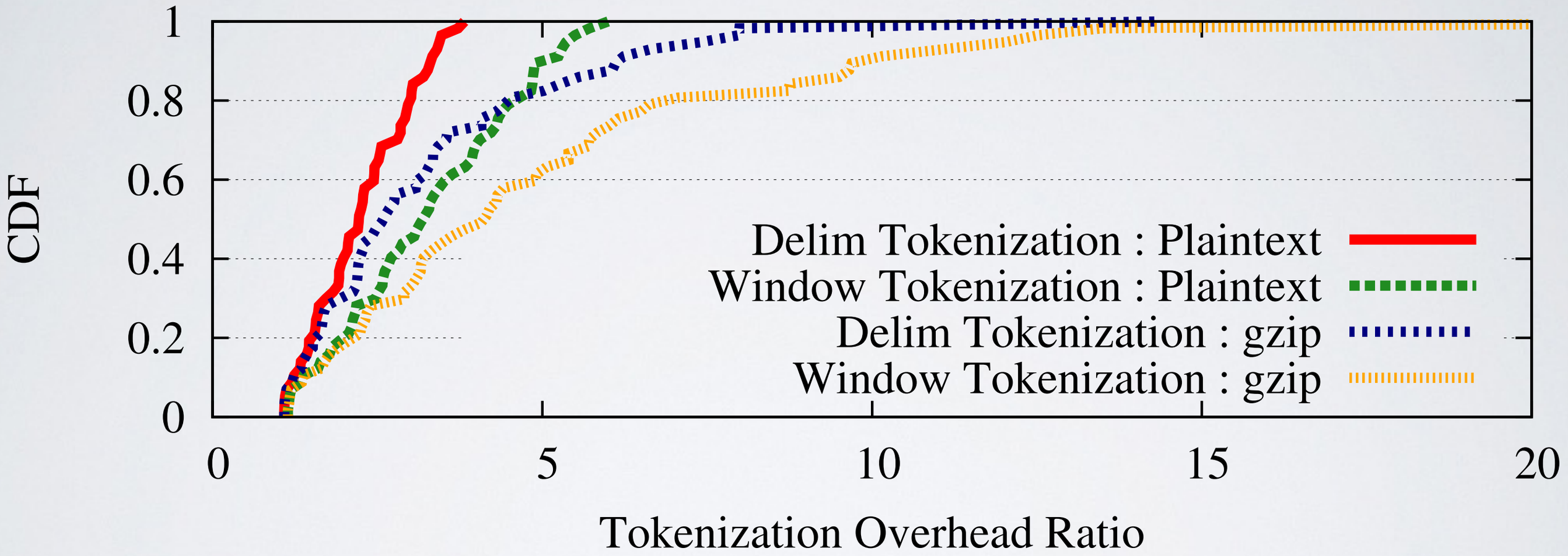
- Alice is a university student connecting her laptop to the campus network.
- Campus policy requires that all traffic be monitored by an IDS to prevent botnet and malware activity from spreading at the university.
- Alice likes the idea of having her laptop protected by these mechanisms, but she is worried by the idea of someone being able to read her traffic and private Facebook messages.

Usage Scenario #2

Bob

- Bob is a father with two small children at home.
- His ISP offers a parental filtering service to block access to pornography.
- Bob would like to opt-in to this service.
- However, Bob read a news article about ISPs selling user data to marketers, and does not want to allow his ISP read all his traffic and sell it to marketers.

Bandwidth Overhead from Tokens



Many pages are gzipped; encrypted data cannot be compressed.