Technologie anonimowości w komunikacji elektronicznej

Mirosław Kutyłowski

Zjazd PTM 2004

Prywatność komunikacji

- szyfrowanie danych
- podpisy cyfrowe
- ...

Prywatność komunikacji

- szyfrowanie danych
- podpisy cyfrowe
- ...
- ale problemy z ukryciem komunikacji

Systemy komunikacji elektronicznej

- sieci TCP/IP – wiadomości dzielone na pakiety, pakiety przesyłane w sposób niekontrolowalny przez nadawcę
- wiele serwerów po drodze ma pełny dostęp do wiadomości
- droga nieznana
- adres docelowy MUSI być widoczny, *adres nadawcy może być sfałszowany*

Potrzeba anonimowości komunikacji

- kontakty business to business (*rozmowy w sprawie budowy fabryki samochodów...*)

Potrzeba anonimowości komunikacji

- kontakty business to business (*rozmowy w sprawie budowy fabryki samochodów...*)
- ochrona konsumenta (*spam reklamowy*)

Potrzeba anonimowości komunikacji

- kontakty business to business (*rozmowy w sprawie budowy fabryki samochodów...*)
- ochrona konsumenta (*spam reklamowy*)
- ochrona prywatności („*Big Brother*“)

Potrzeba anonimowości komunikacji

- kontakty business to business (*rozmowy w sprawie budowy fabryki samochodów...*)
- ochrona konsumenta (*spam reklamowy*)
- ochrona prywatności („*Big Brother*“)
- wybory elektroniczne

Potrzeba anonimowości komunikacji

- kontakty business to business (*rozmowy w sprawie budowy fabryki samochodów...*)
- ochrona konsumenta (*spam reklamowy*)
- ochrona prywatności (*„Big Brother“*)
- wybory elektroniczne
- bezpieczeństwo ekonomiczne i polityczne kraju

Dzięki komunikacji elektronicznej “białe szpiegostwo“ jest tanie, nie trzeba wysyłać szpiegów, ...

Co może adwersarz

- GSM, GPRS,
  - podsłuchiwanie transmisji – wraz z zawartością! – z linii przesyłowych dostawcy

Co może adwersarz

- GSM, GPRS,
  - podsłuchiwanie transmisji – wraz z zawartością! – z linii przesyłowych dostawcy
  - na poziomie komunikacji (stacja bazowa- telefon) – trudne, wymaga specjalistycznego sprzętu, wysiłku, ...

Co może adwersarz

- GSM, GPRS,
  - podsłuchiwanie transmisji – wraz z zawartością! – z linii przesyłowych dostawcy
  - na poziomie komunikacji (stacja bazowa- telefon) – trudne, wymaga specjalistycznego sprzętu, wysiłku, ...
- UMTS
  - solidny system pod względem bezpieczeństwa

Co może adwersarz

- GSM, GPRS,
  - podsłuchiwanie transmisji – wraz z zawartością! – z linii przesyłowych dostawcy
  - na poziomie komunikacji (stacja bazowa- telefon) – trudne, wymaga specjalistycznego sprzętu, wysiłku, ...
- UMTS
  - solidny system pod względem bezpieczeństwa
- Internet
  - cały ruch przechodzący przez węzły współpracujące z adwersarzem,
  - adresów, numerów sekwencyjnych, ... nie daje się szyfrować
    – potrzebne dodatkowe mechanizmy
  - tunelowanie, IPSec, ...

Czy należy ufać?

- Enigma sprzedawana przez aliantów po wojnie zaprzyjaźnionym krajom III świata
- historia Crypto AG
- fundamentalne błędy w procedurze transportu klucza, IBM 4758

Wymagania wobec protokołów anonimowej komunikacji

- ograniczony wzrost ruchu w sieci
- skalowalność
- dowodliwa anonimowość

Najprostszy protokół: all-to-all

- każdy wysyła zaszyfrowaną wiadomość tak aby wszyscy mogli ją odebrać
- tylko adresat może ją odczytać

Najprostszy protokół: all-to-all

- każdy wysyła zaszyfrowaną wiadomość tak aby wszyscy mogli ją odebrać
- tylko adresat może ją odczytać

- idealna anonimowość
- możliwość implementacji np. w lokalnej sieci (np. w oparciu o Ethernet, token ring)
- ale zupełnie nierealistyczne rozwiązanie w skali globalnej

Idea Mix-ów Chauma

Mix to rodzaj specjalnego serwera:

- pewna liczba wiadomości wchodzi równocześnie do Mix-a
- każda wiadomość jest przekodowywana
- wiadomości są permutowane losowo
- i wychodzą z Mix-a

Idea Mix-ów Chauma

Mix to rodzaj specjalnego serwera:

- pewna liczba wiadomości wchodzi równocześnie do Mix-a
- każda wiadomość jest przekodowywana
- wiadomości są permutowane losowo
- i wychodzą z Mix-a

Podstawowa własność: przy odpowiednim kodowaniu kryptograficznym nie można powiązać ze sobą kodów na wejściu i wyjściu z Mix-a

Idea Mix-ów Chauma

Mix to rodzaj specjalnego serwera:

- pewna liczba wiadomości wchodzi równocześnie do Mix-a
- każda wiadomość jest przekodowywana
- wiadomości są permutowane losowo
- i wychodzą z Mix-a

Podstawowa własność: przy odpowiednim kodowaniu kryptograficznym nie można powiązać ze sobą kodów na wejściu i wyjściu z Mix-a
(przy ograniczonych zasobach obliczeniowych obecnie 2^{80} operacji jest niewykonalne – i to się zbytnio nie zmieni!)

Sieci Mix-ów

- kaskada mixów – każdy mix jest administrowany przez kogo innego
  (zastosowanie: wybory elektroniczne, schemat Chauma)

Sieci Mix-ów

- kaskada mixów – każdy mix jest administrowany przez kogo innego
  (zastosowanie: wybory elektroniczne, schemat Chauma)
- przetwarzanie równoległe – małe mix-y permutują małe zbiory wiadomości, wymieniając sie wynikami, znów permutują, ...

Sieci Mix-ów

- kaskada mixów – każdy mix jest administrowany przez kogo innego
  (zastosowanie: wybory elektroniczne, schemat Chauma)
- przetwarzanie równoległe – małe mix-y permutują małe zbiory wiadomości, wymieniając sie wynikami, znów permutują, ...

Problemy:

- jak łączyć ze sobą mix-y?
- jak blisko jesteśmy generowaniu losowej permutacji?

Zastosowania

Anonymous remailer

- spam
- ukrywanie swej tożsamości na listach dyskusyjnych

Zastosowania

Anonymous remailer

- spam
- ukrywanie swej tożsamości na listach dyskusyjnych

- ale też w szlachetnych celach:
  pokazywanie słabości w mechanizmach bezpieczeństwa
  (systemy operacyjne, algorytmy szyfrujące,...),

Dining Cryptographers

- Alicja albo Bob chce mi wysłać pojedynczy bit, ale tak abym nie wiedział od kogo go dostałem,

Dining Cryptographers

- Alicja albo Bob chce mi wysłać pojedynczy bit, ale tak abym nie wiedział od kogo go dostałem,
- Alicja i Bob rzucają monetą, niech wynik będzie równy *b*

Dining Cryptographers

- Alicja albo Bob chce mi wysłać pojedynczy bit, ale tak abym nie wiedział od kogo go dostałem,
- Alicja i Bob rzucają monetą, niech wynik będzie równy b
- jeśli osoba X nie chce przekazać mi bitu, to wysyła mi b,

Dining Cryptographers

- Alicja albo Bob chce mi wysłać pojedynczy bit, ale tak abym nie wiedział od kogo go dostałem,
- Alicja i Bob rzucają monetą, niech wynik będzie równy b
- jeśli osoba X nie chce przekazać mi bitu, to wysyła mi b,
- jeśli osoba X chce przekazać mi o, to wysyła b, a jeśli 1, to wysyła mi $1 - b$,

Dining Cryptographers

- Alicja albo Bob chce mi wysłać pojedynczy bit, ale tak abym nie wiedział od kogo go dostałem,
- Alicja i Bob rzucają monetą, niech wynik będzie równy b
- jeśli osoba X nie chce przekazać mi bitu, to wysyła mi b,
- jeśli osoba X chce przekazać mi o, to wysyła b, a jeśli 1, to wysyła mi $1 - b$,
- dekodowanie: XOR na otrzymanych bitach
  - jeśli było wysyłane 0: $b \text{ XOR } b = 0$
  - jeśli było wysyłane 1: $b \text{ XOR } (1 - b) = 1$

Dining Cryptographers

- *dekodowanie: XOR na otrzymanych bitach*
  - *jeśli było wysyłane* 0*:* $b \; XOR \; b = 0$
  - *jeśli było wysyłane* 1*:* $b \; XOR \; (1 - b) = 1$
- anonimowość:
  - jeśli było wysyłane 0: Alicja i Bob przesłali mi to samo
  - jeśli było wysyłane 1: Alicja i Bob przesłali różne bity, aby wiedzieć kto wysłał b a kto $1 - b$ trzeba znać losowe b

Dining Cryptographers

- *dekodowanie: XOR na otrzymanych bitach*
  - *jeśli było wysyłane* 0*:* $b \; XOR \; b = 0$
  - *jeśli było wysyłane* 1*:* $b \; XOR \; (1 - b) = 1$
- anonimowość:
  - jeśli było wysyłane 0: Alicja i Bob przesłali mi to samo
  - jeśli było wysyłane 1: Alicja i Bob przesłali różne bity, aby wiedzieć kto wysłał b a kto $1 - b$ trzeba znać losowe b
- idealna anonimowość
- problemy ze skalowalnością

Bulletin Board

- wspólny kanał komunikacyjny (np. transmisja z satelity)
- wiadomości szyfrowane
- każdy może odebrać zaszyfrowaną wiadomość, ale odczytać może tylko osoba posiadająca odpowiedni klucz

Cebulki

- nadawca wiadomości wybiera losową ścieżkę, po jakiej wiadomość ma iść do odbiorcy
- każdy serwer na ścieżce pozna tylko:
  - od kogo dostał wiadomość
  - i komu ją dalej przekazał
- odczytanie innych informacji z cebulki (adres końcowy, nadawca, ...) ma być niewykonalne przy ograniczonych mocach obliczeniowych

Konstrukcja cebulki

A wysyła wiadomość *m* zakodowaną jako *O* do serwera *B*:

- *A* wybiera losowo λ pośrednich węzłów $J_1, \ldots, J_\lambda$;

Konstrukcja cebulki

A wysyła wiadomość m zakodowaną jako O do serwera B:

- A wybiera losowo λ pośrednich węzłów $J_1, \ldots, J_\lambda$;
- A tworzy cebulkę:
  $O :=$

$$\mathrm{Enc}_B(m)$$

Konstrukcja cebulki

A wysyła wiadomość m zakodowaną jako O do serwera B:

- A wybiera losowo λ pośrednich węzłów $J_1, \ldots, J_\lambda$;
- A tworzy cebulkę:
  $O :=$
  $$\mathrm{Enc}_{J_\lambda}(\mathrm{Enc}_B(m), B)$$

Konstrukcja cebulki

A wysyła wiadomość m zakodowaną jako O do serwera B:

- A wybiera losowo λ pośrednich węzłów $J_1, \ldots, J_\lambda$;
- A tworzy cebulkę:
  $O :=$

$$\mathrm{Enc}_{J_{\lambda-1}}(\mathrm{Enc}_{J_\lambda}(\mathrm{Enc}_B(m), B), J_\lambda)$$

Konstrukcja cebulki

A wysyła wiadomość m zakodowaną jako O do serwera B:

- A wybiera losowo λ pośrednich węzłów $J_1, \ldots, J_\lambda$;
- A tworzy cebulkę :
  $O :=$
  $\mathrm{Enc}_{J_1}(\ldots(\mathrm{Enc}_{J_{\lambda-1}}(\mathrm{Enc}_{J_\lambda}(\mathrm{Enc}_B(m), B), J_\lambda), J_{\lambda-1})\ldots, J_2)$.

Przetwarzanie cebulek

A wysyła wiadomość *m* zakodowaną jako *O* do serwera *B*:

- *A* przesyła *O* do J_1

Przetwarzanie cebulek

A wysyła wiadomość m zakodowaną jako O do serwera B:

- A przesyła O do J_1
- J_1 deszyfruje O i otrzymuje pewne (O', J_2)

Przetwarzanie cebulek

A wysyła wiadomość m zakodowaną jako O do serwera B:

- A przesyła O do J_1
- J_1 deszyfruje O i otrzymuje pewne (O', J_2)
- J_1 przesyła O' do J_2

Przetwarzanie cebulek

A wysyła wiadomość m zakodowaną jako O do serwera B:

- A przesyła O do J_1
- J_1 deszyfruje O i otrzymuje pewne (O', J_2)
- J_1 przesyła O' do J_2
- J_2 deszyfruje ..
- J_2 przesyła .. do J_3

Przetwarzanie cebulek

A wysyła wiadomość m zakodowaną jako O do serwera B:

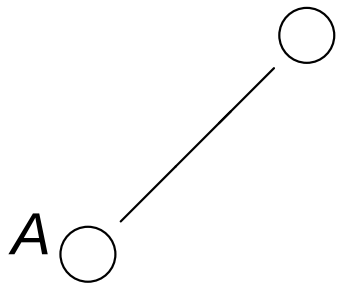
- A przesyła O do J_1
- J_1 deszyfruje O i otrzymuje pewne (O', J_2)
- J_1 przesyła O' do J_2
- J_2 deszyfruje ..
- J_2 przesyła .. do J_3
- ...

Protokół cebulkowy w akcji

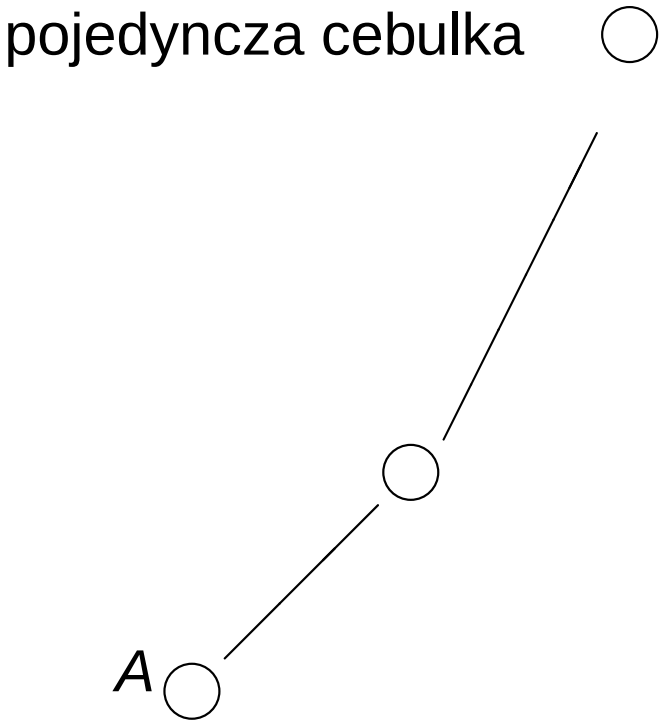
pojedyncza cebulka

Protokół cebulkowy w akcji

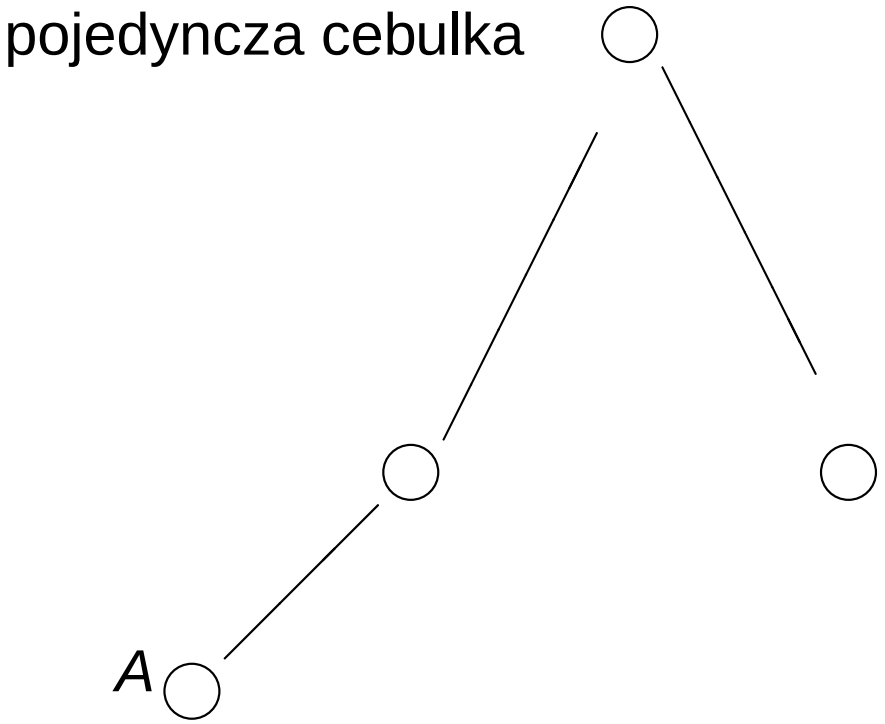
pojedyncza cebulka



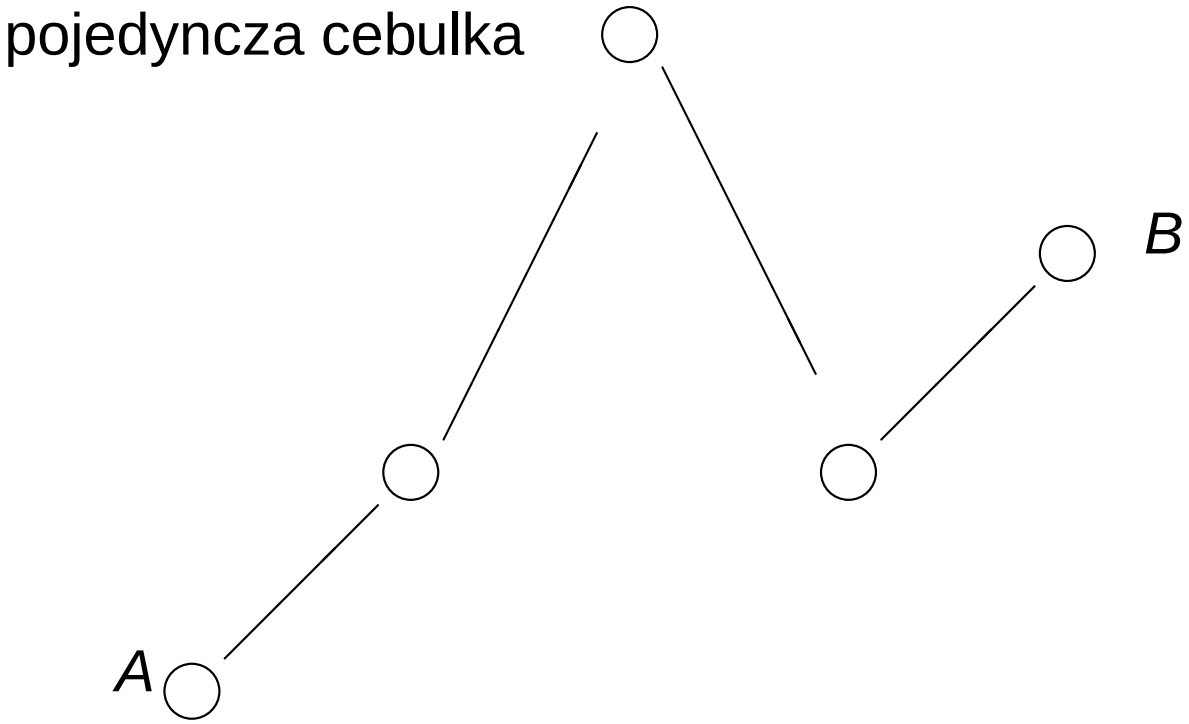
Protokół cebulkowy w akcji



Protokół cebulkowy w akcji

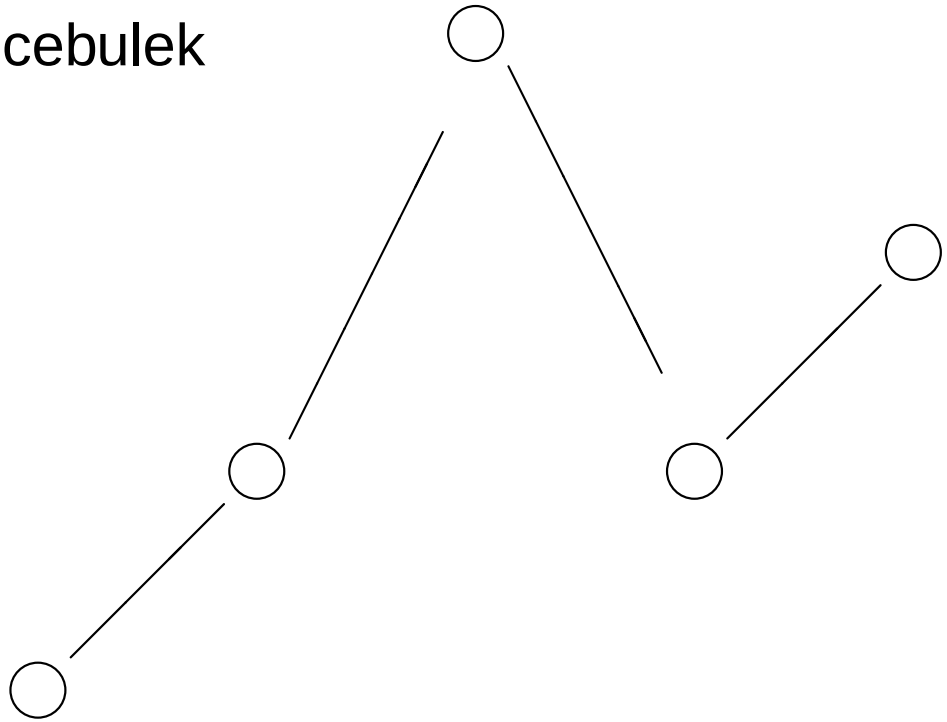


Protokół cebulkowy w akcji

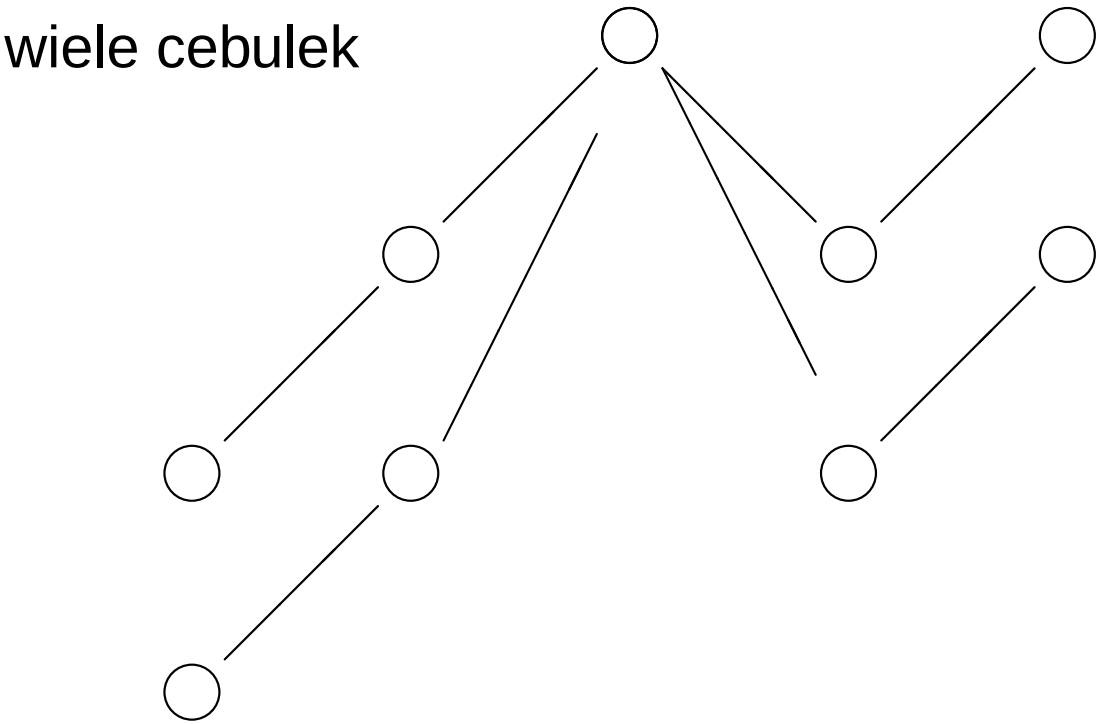


Protokół cebulkowy w akcji

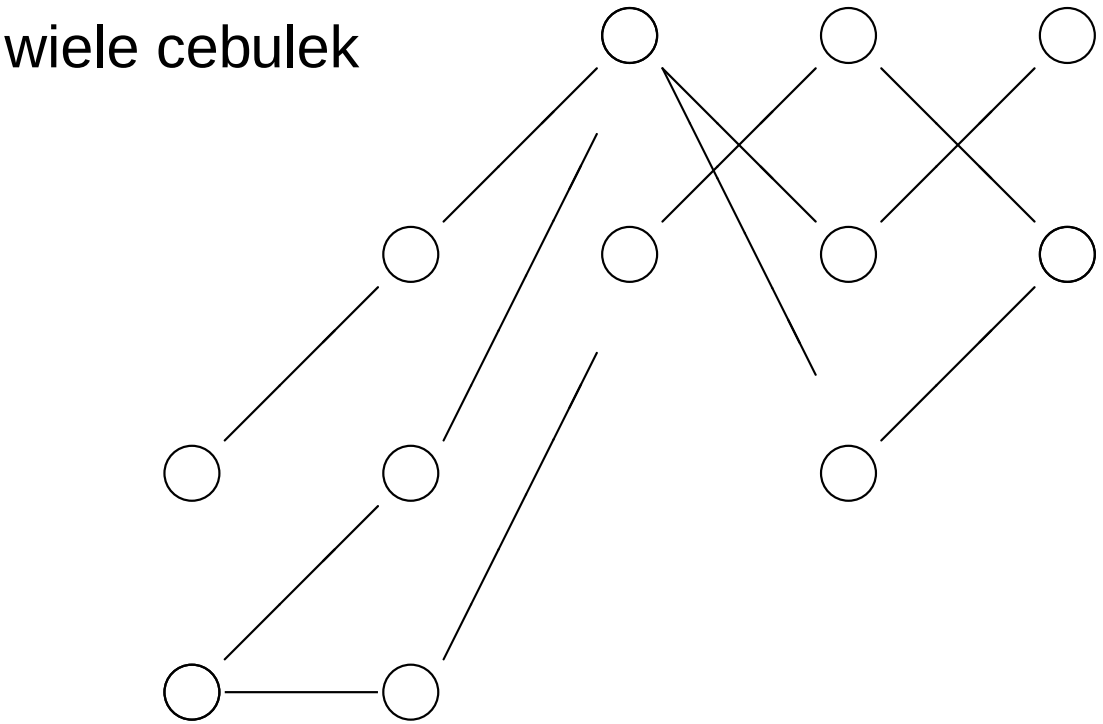
wiele cebulek



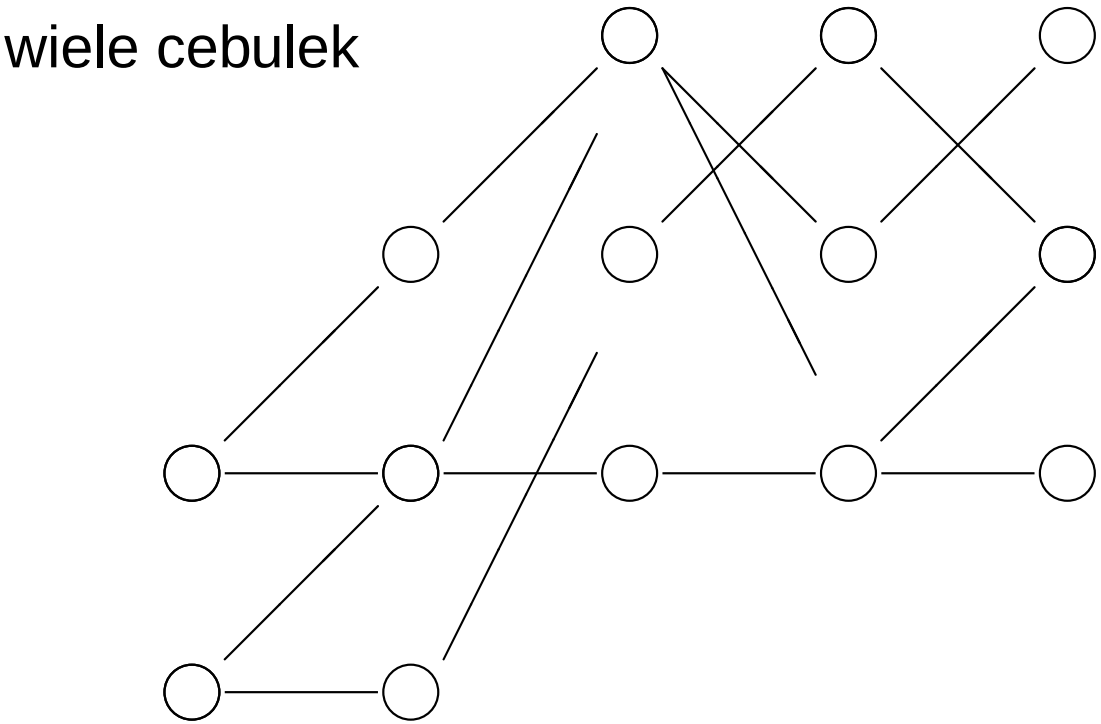
Protokół cebulkowy w akcji



Protokół cebulkowy w akcji

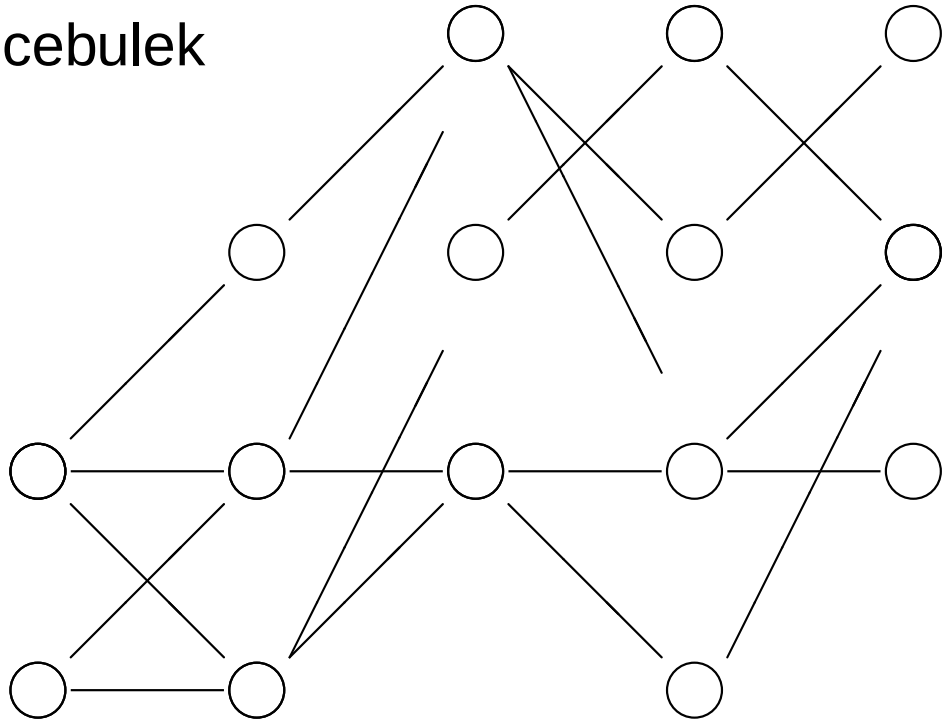


Protokół cebulkowy w akcji

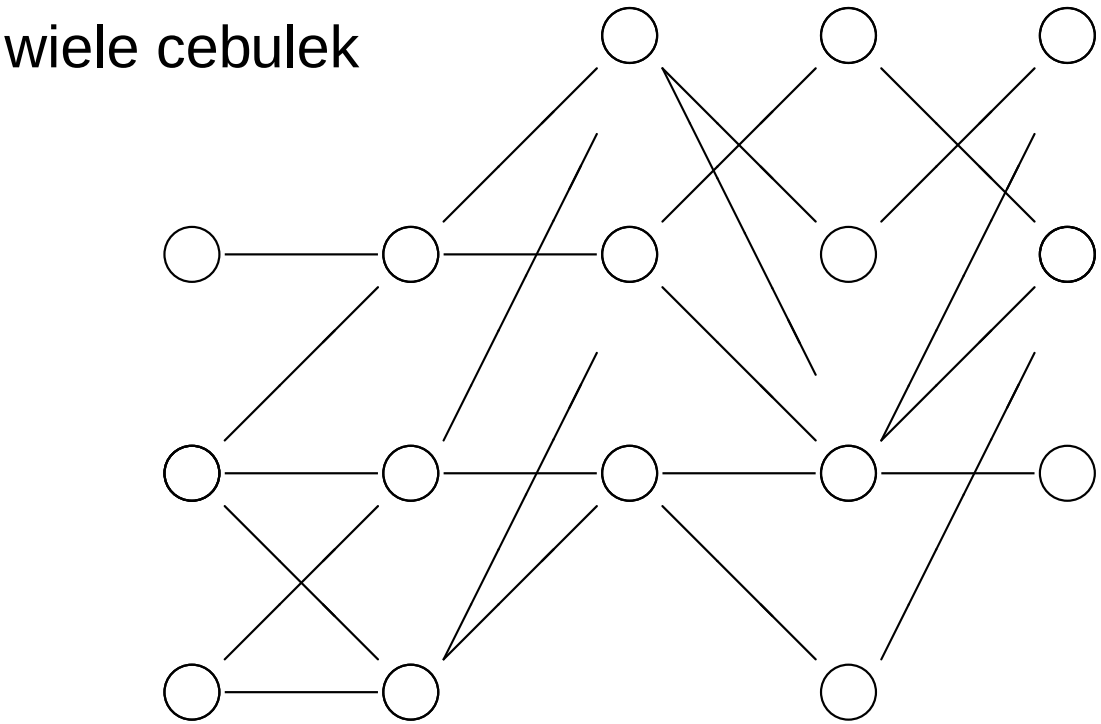


Protokół cebulkowy w akcji

wiele cebulek

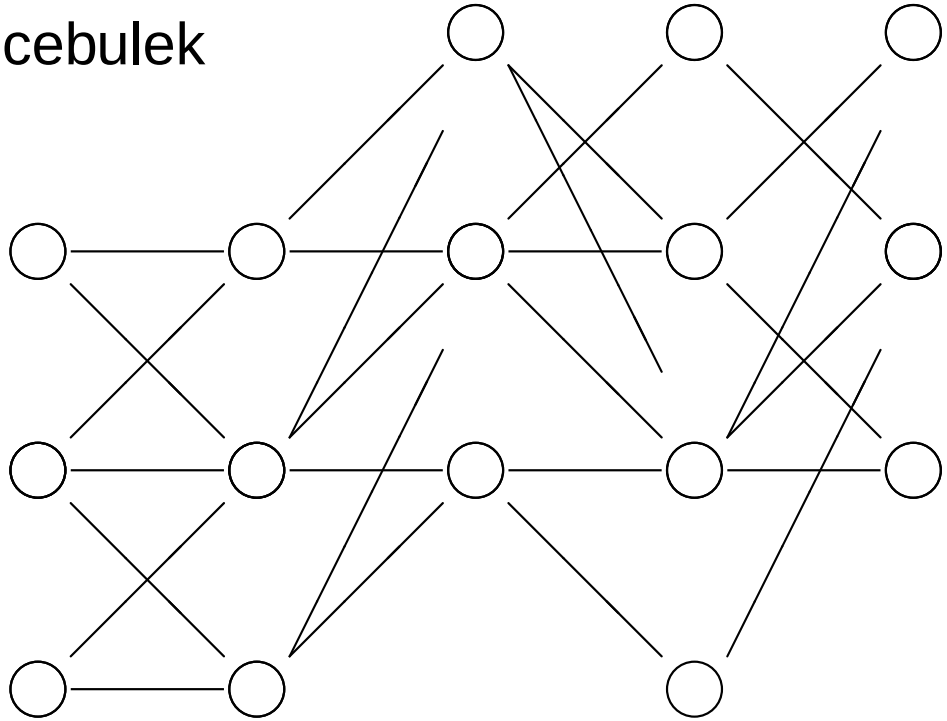


Protokół cebulkowy w akcji

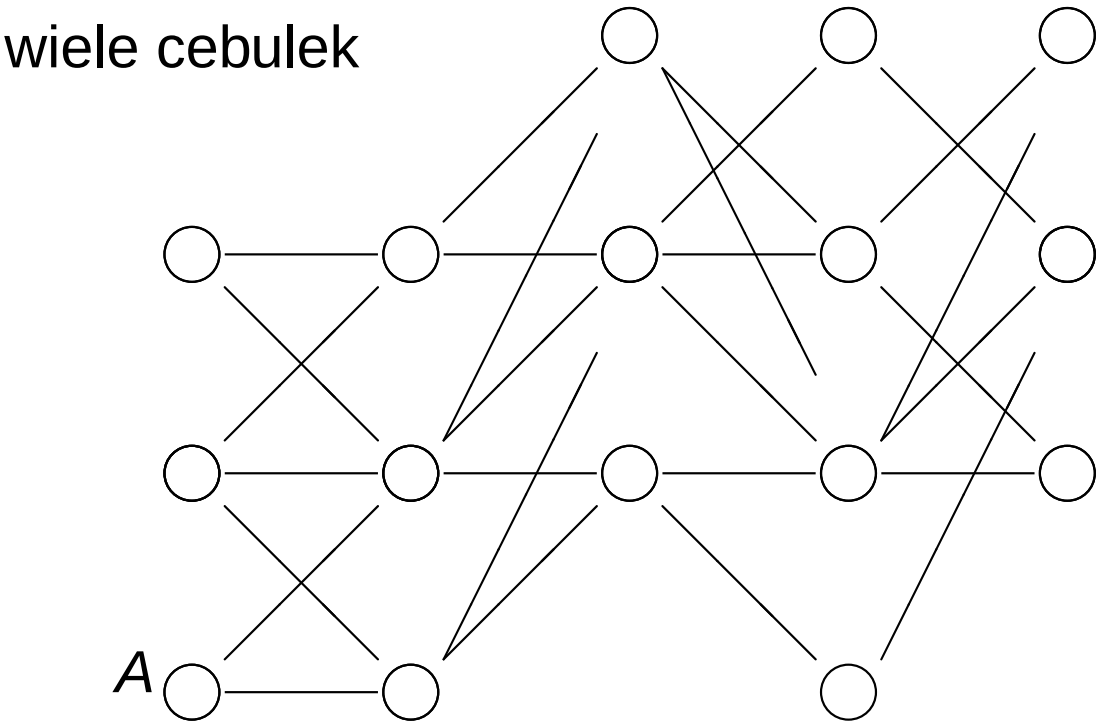


Protokół cebulkowy w akcji

wiele cebulek



Protokół cebulkowy w akcji



miejsce przeznaczenia wiadomości wysłanej przez *A*?

Anonimowość

Co oznacza “anonimowość”?

- “nie można określić dokąd Alicja przesłała wiadomość”

Anonimowość

Co oznacza “anonimowość”?

- “nie można określić dokąd Alicja przesłała wiadomość”
  CZY

Anonimowość

Co oznacza “anonimowość”?

- “nie można określić dokąd Alicja przesłała wiadomość” CZY
- żadne istotne informacje odnośnie zachowania uczestników protokołu komunikacyjnego nie są możliwe do wydedukowania

Dlaczego ścisła, matematyczna definicja anonimowości jest ważna

Przykład: schemat wyborów drogą elektroniczną

- Ewa analizuje prawdopodobieństwo, że Alicja głosowała na partię Y, dla każdego Y,
- jeśli rozkład zbliżony do jednostajnego, to często uważa się, że dowiedziono, iż schemat wyborów jest bezpieczny

Dlaczego ścisła, matematyczna definicja anonimowości jest ważna

Przykład: schemat wyborów drogą elektroniczną

- Ewa analizuje prawdopodobieństwo, że Alicja głosowała na partię Y, dla każdego Y,
- jeśli rozkład zbliżony do jednostajnego, to często uważa się, że dowiedziono, iż schemat wyborów jest bezpieczny

NIEPRAWDA!

Dlaczego ścisła, matematyczna definicja anonimowości jest ważna

- Ewa być może nie potrafi określić na kogo głosowała Alicja,

Dlaczego ścisła, matematyczna definicja anonimowości jest ważna

- Ewa być może nie potrafi określić na kogo głosowała Alicja,
- ale może wykazać, że Alicja i Jurek głosowali na tę samą partię z prawdopodobieństwem 99%,

Dlaczego ścisła, matematyczna definicja anonimowości jest ważna

- Ewa być może nie potrafi określić na kogo głosowała Alicja,
- ale może wykazać, że Alicja i Jurek głosowali na tę samą partię z prawdopodobieństwem 99%,
- ... i wystarczy kupić od Jurka informacje na temat jego głosu.

Prace na temat anonimowości

Ron Berman, Amos Fiat, Amnon Ta-Shma powiedzieli:

- *Literally dozens (hundreds?) of papers since, dedicated conferences, etc., etc.*
- *Many implementations*
- *Typical paper:*
  Attack on prior protocol(s)
  Suggest new protocol
  Repeat
- *Very few attempts to give rigorous definitions, let alone proofs*
- *Notable exception: Rackoff and Simon, 1993*

k-anonymity

- pojęcie używane intensywnie w systemach bazodanowych,
- każda osoba ma być nierozróżnialna od *k* innych osób
  jeśli prokurator oskarża Alicję o pewien czyn, to Alicja broni się wskazując k innych, równo prawdopodobnych podejrzanych

k-anonymity

- pojęcie używane intensywnie w systemach bazodanowych,
- każda osoba ma być nierozróżnialna od *k* innych osób
  jeśli prokurator oskarża Alicję o pewien czyn, to Alicja broni się wskazując k innych, równo prawdopodobnych podejrzanych

Problem:

- niski poziom anonimowości
- skuteczny przy unikaniu odpowiedzialności przy domniemaniu niewinności, ale nieskuteczny przy domniemaniu winy

Anonymity set

- niech *A* będzie zbiorem wszystkich potencjalnych odbiorców wiadomości *m* (tj. takich, dla których prawdopodobieństwo otrzymania *m* jest dodatnie - przestrzeń jest dyskretna)
- *A* jest nazywany *anonymity set* dla wiadomości *m*
- miara anonimowości: **moc zbioru** *A*

Anonymity set

- niech *A* będzie zbiorem wszystkich potencjalnych odbiorców wiadomości *m* (tj. takich, dla których prawdopodobieństwo otrzymania *m* jest dodatnie - przestrzeń jest dyskretna)
- *A* jest nazywany *anonymity set* dla wiadomości *m*
- miara anonimowości: **moc zbioru** *A*

Problem:

- jeśli moc zbioru jest mała, to mamy niski poziom anonimowości,
- jeśli moc jest duża, to niekoniecznie świadczy to o wysokim poziomie anonimowości,

Najwyższe prawdopodobieństwa

- miara anonimowości: **najwyższe prawdopodobieństwo w *anonymity set***

Entropia

- rozważmy prawdopodobieństwa dla każdej lokalizacji z A
- miara anonimowości: **entropia tego rozkładu**
- motywacja: entropia mówi, ile średnio bitów potrzeba do określenia aktualnej lokalizacji

Problemy z tymi definicjami

- tylko jedna wiadomość brana pod uwagę
- a zależności między wiadomościami mogą odgrywać kluczową rolę

Problemy z tymi definicjami

oczywisty przykład: “pseudo-mix”

- input: n wiadomości na pozycjach od 0 do $n-1$

Problemy z tymi definicjami

oczywisty przykład: “pseudo-mix”

- input: n wiadomości na pozycjach od 0 do $n - 1$
- przekodowywanie wiadomości metodami kryptograficznymi – jak zwykle

Problemy z tymi definicjami

oczywisty przykład: “pseudo-mix”

- input: n wiadomości na pozycjach od 0 do $n-1$
- przekodowywanie wiadomości metodami kryptograficznymi – jak zwykle
- “permutowanie”:
  - $r < n$ wybierane losowo z jednostajnym prawdopodobieństwem,

Problemy z tymi definicjami

oczywisty przykład: “pseudo-mix”

- input: n wiadomości na pozycjach od 0 do $n-1$
- przekodowywanie wiadomości metodami kryptograficznymi – jak zwykle
- “permutowanie”:
  - $r < n$ wybierane losowo z jednostajnym prawdopodobieństwem,
  - przekodowana wiadomość z pozycji i przesuwana na pozycję $i + r \mod n$.

Problemy z tymi definicjami

oczywisty przykład: “pseudo-mix”

- input: n wiadomości na pozycjach od 0 do $n-1$
- przekodowywanie wiadomości metodami kryptograficznymi – jak zwykle
- “permutowanie”:
  - $r < n$ wybierane losowo z jednostajnym prawdopodobieństwem,
  - przekodowana wiadomość z pozycji i przesuwana na pozycję $i + r \mod n$.
- adwersarz może przesłać swoją wiadomość, zobaczyć gdzie trafiła
  .. i ma r. Pożegnanie z anonimowością!

Problemy z tymi definicjami

oczywisty przykład: “pseudo-mix”

- input: n wiadomości na pozycjach od 0 do $n-1$
- przekodowywanie wiadomości metodami kryptograficznymi – jak zwykle
- “permutowanie”:
  - $r < n$ wybierane losowo z jednostajnym prawdopodobieństwem,
  - przekodowana wiadomość z pozycji i przesuwana na pozycję $i + r \mod n$.
- adwersarz może przesłać swoją wiadomość, zobaczyć gdzie trafiła
  .. i ma r. Pożegnanie z anonimowością!
- ale entropia dla każdej pojedynczej wiadomości wynosi $\log n$ (maksymalna wartość)

Analiza ruchu

rozważmy sieć komunikacyjną z protokołem, w którym wiadomości są kodowane i przekodowywane w sposób nie do złamania

Analiza ruchu

rozważmy sieć komunikacyjną z protokołem, w którym wiadomości są kodowane i przekodowywane w sposób nie do złamania

- ile zyskuje adwersarz przez obserwowanie ruchu?

Analiza ruchu

rozważmy sieć komunikacyjną z protokołem, w którym wiadomości są kodowane i przekodowywane w sposób nie do złamania

- ile zyskuje adwersarz przez obserwowanie ruchu?
- czasami adwersarz dowie się wszystkiego (np. gdy różne wiadomości idą rozłącznymi ścieżkami)

Analiza ruchu

rozważmy sieć komunikacyjną z protokołem, w którym wiadomości są kodowane i przekodowywane w sposób nie do złamania

- ile zyskuje adwersarz przez obserwowanie ruchu?
- czasami adwersarz dowie się wszystkiego (np. gdy różne wiadomości idą rozłącznymi ścieżkami)
- często źródła wiadomości i ich miejsca przeznaczenia nie mogą być ukryte – ale powiązanie ich **ma być trudne**

Punkt widzenia adwersarza bez informacji o ruchu

- dla każdego węzła wie:
  - ile wiadomości wysłał początkowo,
  - ile wiadomości w końcu dostarczono.

Punkt widzenia adwersarza bez informacji o ruchu

- dla każdego węzła wie:
  - ile wiadomości wysłał początkowo,
  - ile wiadomości w końcu dostarczono.
- zmienna losowa π:
  $\pi(j) = i \Leftrightarrow$ wiadomość i-ta dociera do j-tego miejsca przeznaczenia
- rozkład prawdopodobieństwa π określa informacje potencjalnie dostępne adwersarzowi

Punkt widzenia adwersarza z informacją o ruchu

- tak jak poprzednio, ale dodatkowo adwersarz wie, kiedy i po których liniach były przesyłane wiadomości

Rozkład prawdopodobieństwa

- teraz prawdopodobieństwa warunkowe:

$$\Pr[\pi|c]$$

  gdzie c opisuje zaobserwowany ruch
- różne c mogą różnie wpływać na wartości prawdopodobieństw warunkowych,

Rozkład prawdopodobieństwa

- teraz prawdopodobieństwa warunkowe:

$$\Pr[\pi|c]$$

  gdzie c opisuje zaobserwowany ruch
- różne c mogą różnie wpływać na wartości prawdopodobieństw warunkowych,
- cel: prawdopodobieństwa warunkowe powinny być prawie takie same jak te bez informacji o ruchu

Rozkład prawdopodobieństwa

- teraz prawdopodobieństwa warunkowe:

$$\Pr[\pi|c]$$

  gdzie c opisuje zaobserwowany ruch
- różne c mogą różnie wpływać na wartości prawdopodobieństw warunkowych,
- cel: prawdopodobieństwa warunkowe powinny być prawie takie same jak te bez informacji o ruchu
- nie zawsze możliwe
- zmodyfikowany cel: otrzymać tę własność dla prawie wszystkich c, tj. z wysokim prawdopodobieństwem.

Odległość rozkładów

Variation distance

- rozważamy rozkłady μ_1 and μ_2 na przestrzeni dyskretnej Ω
- variation distance:

$$\|\mu_1 - \mu_2\| = \frac{1}{2} \sum_{\omega \in \Omega} |\mu_1(\omega) - \mu_2(\omega)| \ .$$

Definicja anonimowości oparta o variation distance

$$\|\Pi - \Pi|C\| \leq \ldots(\text{jakaś mała liczba})$$

gdzie Π jest rozkładem prawdopodobieństwa π,
$\Pi|C$ jest rozkładem prawdopodobieństwa π pod warunkiem informacji o ruchu C

Definicja anonimowości oparta o *wzajemną informację*

- podejście teorio-informacyjne
- intuicyjnie: ile bitów informacji o Π otrzymujemy poprzez wiedzę o C?

Definicja anonimowości oparta o *wzajemną informację*

- podejście teorio-informacyjne
- intuicyjnie: ile bitów informacji o Π otrzymujemy poprzez wiedzę o C?
- definicje te są z grubsza równoważne – oszacowania Berman, Fiat, Ta-Shma, 2004

Rezultaty na temat protokołu cebulkowego

Jaka długość ścieżki λ gwarantuje wysoką anonimowość (małą *variation distance*)?

Gdy adwersarz widzi cały ruch:

Rezultaty na temat protokołu cebulkowego

Jaka długość ścieżki λ gwarantuje wysoką anonimowość (małą *variation distance*)?

Gdy adwersarz widzi cały ruch:

- Rackoff, Simon (ACM STOC'93):
  bezpieczeństwo dla $\lambda \approx \log^{11} n$,
  specjalne założenie: a i-tym etapie wiadomości pozostają w grupach złożonych z 2^i węzłów, każdy serwer wysyła jedną wiadomość

Rezultaty na temat protokołu cebulkowego

Jaka długość ścieżki λ gwarantuje wysoką anonimowość (małą *variation distance*)?

Gdy adwersarz widzi cały ruch:

- Rackoff, Simon (ACM STOC'93):
  bezpieczeństwo dla $\lambda \approx \log^{11} n$,
  specjalne założenie: a i-tym etapie wiadomości pozostają w grupach złożonych z 2^i węzłów, każdy serwer wysyła jedną wiadomość
- Czumaj, Kanarek, Kutyłowski, Loryś (ACM SODA'99):
  przy tych samych założeniach dla $\lambda \approx \log^2 n$

Rezultaty na temat protokołu cebulkowego

Gdy adwersarz widzi tylko część połączeń:

Rezultaty na temat protokołu cebulkowego

Gdy adwersarz widzi tylko część połączeń:

- Berman, Fiat, Ta-Shma (FC'2004) – dla $\lambda \approx \log^4 n$ dla n wiadomości oraz variation distance $1/n$

Rezultaty na temat protokołu cebulkowego

Gdy adwersarz widzi tylko część połączeń:

- Berman, Fiat, Ta-Shma (FC'2004) – dla $\lambda \approx \log^4 n$ dla n wiadomości oraz variation distance $1/n$
- Gomułkiewicz, Klonowski, Kutyłowski (ISC'2004) – $\lambda \approx \log n$ steps,
  optymalny rezultat — dla $\lambda = o(\log n)$ nie jest to możliwe

Techniki dowodowe - rapid mixing i anonimowość

rozważamy proces stochastyczny przekazywania wiadomości

- w każdym kroku wiadomości są przekodowywane w węzłach i ..
- przesyłane dalej do losowo wybranych lokalizacji

Techniki dowodowe - rapid mixing i anonimowość

rozważamy proces stochastyczny przekazywania wiadomości

- w każdym kroku wiadomości są przekodowywane w węzłach i ..
- przesyłane dalej do losowo wybranych lokalizacji
- adwersarz widzi ruch (cały lub na określonych liniach komunikacyjnych)

Techniki dowodowe - rapid mixing i anonimowość

rozważamy proces stochastyczny przekazywania wiadomości

- w każdym kroku wiadomości są przekodowywane w węzłach i ..
- przesyłane dalej do losowo wybranych lokalizacji
- adwersarz widzi ruch (cały lub na określonych liniach komunikacyjnych)

Jak długo proces musi trwać, aby zbliżyć się do rozkładu stacjonarnego?

Rapid mixing

- nie chodzi o zbieżność do rozkładu stacjonarnego ale o tempo zbieżności
- różnorodne techniki:
  - metoda potencjału
  - coupling
  - path coupling
  - delayed path coupling
  - ...
- $\Leftrightarrow$ algorytmy oparte o rapid mixing

Konkluzje

- istnieją techniki dobrze chroniące anonimowość,
- ale uwaga - diabeł tkwi w szczegółach!

Konkluzje

- istnieją techniki dobrze chroniące anonimowość,
- ale uwaga - diabeł tkwi w szczegółach!
- znaczący postęp w matematycznych metodach oceny poziomu anonimowości
- ale też rosnące szpiegostwo elektroniczne

Dziekuję za uwagę!