

이 발명을 지원한 국가연구개발사업

과제고유번호 C1090-0801-0016

부처명 지식경제부

연구사업명 대학 IT연구센터 육성 지원사업

연구과제명 해킹·바이러스 대응기술 연구

주관기관 충남대학교

연구기간 2003. 08. 01~2011. 12. 31

특허청구의 범위

청구항 1

하나 또는 둘 이상의 패턴으로 이루어진 패턴집합을 이용한 패턴매칭 방법으로서,

(A) 하기의 소단계에 의해 부분패턴집합을 결정하는 단계:

(a) 패턴집합의 최소길이 패턴에 대하여 2이상 최소패턴의 길이[ℓ] 이하의 임의의 크기로 부분패턴의 크기 $[b](2 \leq b \leq \ell)$ 를 결정하는 소단계;

(b) 패턴집합의 모든 패턴에 대하여, 부분패턴의 수 $[n]$ 만큼 연속된 길이 b 인 부분패턴을 추출하는 소단계;

(c) 상기 추출된 각 부분패턴에 대한 모패턴의 길이 $[L]$ 및 모패턴에서의 부분패턴의 위치정보를 포함하는 메타정보로 이루어진 부분패턴 테이블을 구성하는 소단계;

(B) 하기의 소단계에 의해 페이로드 또는 문자열이 상기 패턴집합을 이루는 패턴 중 최소한 하나의 패턴을 포함하고 있는지를 검색하는 단계:

(a) 상기 페이로드 또는 문자열의 b 개의 연속 문자열로 된 비교부분과 상기 부분패턴집합의 부분패턴들과의 일치여부를 비교하는 소단계;

(b) ① 일치하는 부분패턴이 존재하지 않으면 문자열을 다음으로 n 만큼 이동하여 상기 (Ba)단계를 수행하며, ② 일치하는 부분패턴이 존재하면, 상기 부분패턴 테이블을 참조하여 일치 부분패턴의 모패턴이 상기 페이로드 또는 문자열과 일치하는지 확인하여 일치하지 않으면 문자열을 하부로 n 만큼 이동하여 상기 (Ba)단계를 수행하며, 일치하면 '패턴매칭 성공'이라는 결과를 보고하는 소단계;

(c) 상기 페이로드 또는 문자열의 마지막 b 개 문자열까지 비교결과 패턴매칭이 확인되지 않으면 '패턴매칭 실패'라는 결과를 보고하는 소단계;

를 포함하는 것을 특징으로 하는 여러 개의 연속된 부분패턴을 이용한 다중 패턴매칭 방법.

청구항 2

제 1 항에 있어서,

상기 패턴에서 추출한 여러 개의 연속된 부분패턴에서 중복된 것을 제거하고 부분패턴에 대해 상기 부분패턴 테이블에 각각의 메타정보를 저장하는 것을 특징으로 하는 여러 개의 연속된 부분패턴을 이용한 다중 패턴매칭 방법.

청구항 3

제 1 항 또는 제 2 항에 있어서,

상기 페이로드와 동일한 플로우에 속하는 후속 페이로드가 있는 경우 전단의 페이로드 마지막과 후단의 페이로드 처음을 연결하여 패턴매칭을 검색하는 것을 특징으로 하는 여러 개의 연속된 부분패턴을 이용한 다중 패턴매칭 방법.

청구항 4

제 1 항 또는 제 2 항에 있어서,

상기 단계 (A)에는, (d) 추출된 부분패턴에 대한 해시값을 정하는 소단계;를 추가로 포함하며,

상기 단계 (B)의 (a)에서는 비교부분에 대한 해시값을 구하고 상기 부분패턴의 해시값들과의 일치여부를 비교하는 것을 특징으로 하는 여러 개의 연속된 부분패턴을 이용한 다중 패턴매칭 방법.

명세서

발명의 상세한 설명

기술 분야

- <1> 본 발명은 침입탐지 또는 차단시스템의 요소 기술이 되는 패킷의 페이로드 부분에서 규칙집합에 존재하는 문자열을 검색하는 패턴매칭을 개선한 방법에 관한 것이다.

배경 기술

- <2> 인터넷이 광범위하게 사용되면서 웜이나 트로이목마, 바이러스, DoS 등과 같은 다양한 형태의 침해가 나타나고 그 파괴효과도 커지고 있다. 따라서 네트워크상의 정보를 보호하기 위한 네트워크 침해 대응 방법 및 기타 관련 장치가 필요하게 되었다. 이를 위한 가장 기본적인 기술이 바로 고속의 패턴매칭 기술이다.
- <3> 패턴매칭은, 긴 텍스트 파일에서 우리가 원하는 글자를 찾는 것부터 시작해서 네트워크 침입 탐지나 바이러스 탐지 등과 같은 보안 시스템, 스팸 메일 제거 시스템, 데이터베이스 시스템 등 다양한 분야에서 다양한 용도로 활용된다. 특히, 패킷의 페이로드 부분에서 특정한 문자열을 검색하는 패턴매칭 방법은 규칙을 이용하여 비정상 행위를 탐지하는 침입탐지 또는 차단시스템의 중요한 요소 기술이다.
- <4> 관리자들은 악의적인 공격으로부터 내부 네트워크를 보호하기 위하여 침입탐지 또는 차단 시스템을 사용하는데, 공격이 다양해지고 그 수가 증가하면서 공격을 탐지하기 위한 패턴매칭의 규칙수도 증가하였다. 규칙의 수가 증가하면서 패턴매칭 비용이 침입 탐지나 차단에서 차지하는 비중이 커졌다.
- <5> 다중 패턴매칭 알고리즘인 Wu-Manber 알고리즘은 문자열을 검색하는 과정에서의 건너뛴, 즉 이동(shift)거리를 증가시킴으로써 패턴매칭 성능을 개선하기 위한 것이다. 종래의 Wu-Manber 알고리즘은 전처리 단계에서 SHIFT 테이블과 HASH 테이블, PREFIX 테이블을 구성하고, 검색단계에서 각 테이블을 이용하여 페이로드에서 패턴을 찾는 방법을 사용한다. 그러나 패턴의 개수가 많아지면 페이로드에서 패턴을 비교하면서 건너뛴 수 있는 이동거리가 평균적으로 감소한다. 또한 SHIFT 테이블과 HASH 테이블, PREFIX 테이블을 비교한 후 문자열을 비교하는 검색단계의 복잡도가 커져서 시간 및 비용을 증가시키는 요인이 되고 있다.

발명의 내용

해결 하고자하는 과제

- <6> 본 발명은 시스템에 대한 공격이 다양화되고 공격 형태가 구조화됨에 따라 규칙의 개수가 지속적으로 증가하면서 상대적으로 많은 오버헤드를 발생시키는 종래의 패턴매칭 방법을 개선하기 위한 새로운 다중 패턴매칭 방법을 제공하고자 한다.

과제 해결수단

- <7> 전술한 과제를 해결하기 위한 본 발명은, 하나 또는 둘 이상의 패턴으로 이루어진 패턴집합을 이용한 패턴매칭 방법으로서, (A) (a) 패턴집합의 최소길이 패턴(최소패턴)에 대하여 2이상 최소패턴의 길이[ℓ] 이하의 임의의 길이로 부분패턴의 길이[b]($2 \leq b \leq \ell$)를 결정하는 소단계; (b) 상기 최소패턴에서 부분패턴을 추출하며, 패턴집합의 모든 패턴의 임의의 위치에서, 상기 최소패턴에서 추출된 부분패턴의 수[n]만큼 연속된 길이[b]의 부분패턴을 추출하는 소단계; (c) 상기 추출된 각 부분패턴에 대한 모패턴의 길이[L] 및 모패턴에서의 부분패턴의 위치정보를 포함하는 메타정보를 포함하는 부분패턴 테이블을 구성하는 소단계;에 의해 부분패턴집합을 결정하는 부분패턴추출단계:와, (B) (a) 페이로드 또는 문자열에서 b개의 연속 문자열로 된 비교부분과 상기 부분패턴집합의 부분패턴들과의 일치여부를 비교하는 소단계; (b) ① 일치하는 부분패턴이 존재하지 않으면 문자열을 다음으로 n만큼 이동하여 상기 (Ba)단계를 수행하며, ② 일치하는 부분패턴이 존재하면, 상기 부분패턴 테이블을 참조하여 일치하는 부분패턴의 모패턴이 상기 페이로드 또는 문자열의 비교부분을 포함하는 영역과 일치하는지 확인하여 일치하지 않으면 문자열을 하부로 n만큼 이동하여 상기 (Ba)단계를 수행하며, 일치하면 '패턴매칭 성'이라는 결과를 보고하는 소단계; (c) 상기 페이로드 또는 문자열의 마지막 b개 문자열까지 비교결과 패턴매칭이 확인되지 않으면 '패턴매칭 실패'라는 결과를 보고하는 소단계;에 의해 페이로드 또는 문자열이 상기 패턴집합을 이루는 패턴 중 최소한 하나의 패턴을 포함하고 있는지를 검색하는 매칭확인단계:를 포함하는 여러 개의 연속된 부분패턴을 이용한 다중 패턴매칭 방법에 관한 것이다.
- <8> 이해의 편의를 위해 주요 용어에 대해 다음 표와 같은 약자를 사용하기로 한다. 본 발명에서 '길이'라 함은 문자열의 문자수일 수도 있고, 바이트수일 수도 있고 비트수일 수도 있다. 이하에서는 문자수 또는 바이트수로 혼용한다.

최소패턴의 길이	ℓ
부분패턴의 길이	b
최소패턴의 부분패턴 수	n
모패턴의 길이	L

<9>

<10> 이때, 상기 패턴에서 추출한 여러 개의 연속된 부분패턴에서 중복된 부분패턴이 존재하는 경우 상기 부분패턴 테이블에 각각의 메타정보를 저장하는 것이 바람직하다.

<11> 물론, 상기 페이로드와 동일한 플로우에 속하는 후속 페이로드가 있는 경우 전단의 페이로드 마지막과 후단의 페이로드 처음을 연결하여 패턴매칭을 검색하는 것이 바람직하다.

<12> 본 발명에서는, 부분패턴의 저장공간을 줄이고, 비교속도를 증대시키기 위해, 상기 단계 (A)에는, (d) 추출된 부분패턴에 대한 해시(hash)값을 정하는 소단계;를 추가로 포함하며, 상기 단계 (B)의 (a)에서는 비교부분에 대한 해시값을 구하고 상기 부분패턴의 해시값들과의 일치여부를 비교하도록 할 수도 있다.

효과

<13> 본 발명은, 여러 개의 연속된 부분패턴을 사용하여 효율적으로 패턴매칭을 수행함으로써 비정상탐지 및 오용탐지 기법을 사용하는 침입탐지시스템에서 패킷의 페이로드 부분에서 침입여부를 판단할 목적으로 문자열을 탐색하는데 활용될 수 있으며, 크기가 큰 문서에서 여러 개의 특정 패턴의 존재 유무를 검사하기 위하여 문자열을 검색할 때도 활용될 수 있다.

발명의 실시를 위한 구체적인 내용

<14> 이하 첨부된 도면을 참조하여 본 발명을 보다 상세히 설명한다. 그러나 이러한 도면은 본 발명의 기술적 사상의 내용과 범위를 쉽게 설명하기 위한 예시일 뿐, 이에 의해 본 발명의 기술적 범위가 한정되거나 변경되는 것은 아니다. 또한 이러한 예시에 기초하여 본 발명의 기술적 사상의 범위 안에서 다양한 변형과 변경이 가능함은 당업자에게는 당연할 것이다.

<15> 각 패턴에 대하여 모패턴의 길이보다 작은 동일한 길이 $[b]$ 의 연속된 부분패턴을 추출할 수 있다. 도 1에 6개의 문자로 구성된 패턴(11)에 존재하는 3바이트 크기의 부분패턴(12)과 4바이트 길이의 부분패턴(13)을 각각 예시적으로 나타내었다. 이와 같이 패턴에서 추출할 수 있는 여러 개의 연속된 부분패턴들을 이용하면 패킷의 페이로드 부분에 패턴이 존재할 경우 패턴의 일치여부의 확인 없이 건너 뛸 수 있는 이동거리를 부분패턴의 수 $[n]$ 만큼 늘리더라도 패턴매칭의 결과가 항상 올바름을 보장할 수 있다.

<16> 예를 들면, 도 2에서와 같이 페이로드 또는 문자열(21)에 도면 1의 패턴(11)이 존재할 경우, 만약 b 가 3바이트인 연속된 부분패턴(12) 4개가 주어진다면 패턴을 포함하는 페이로드 또는 문자열(21)의 처음부터 일치하는 부분패턴이 존재하는지 확인하기 위하여 비교하는 과정에서 패턴(11)의 부분패턴수인 4만큼 건너뛰며 패턴매칭해도 여러 개의 연속된 부분패턴(12) 중에서 하나의 부분패턴(22)을 찾을 수 있다.

<17> 본 발명은 이와 같은 사실에 기초한 것이다.

<18> 이하 부분패턴추출단계와 매칭확인단계로 구분하여 설명한다.

<19> (A) 부분패턴추출단계

<20> 부분패턴을 추출하는 과정을 도 3에 도시하였다.

<21> 패턴집합에서 최소패턴의 길이 $[\ell]$ 를 확인하여 부분패턴의 길이 $[b]$ 를 임의로 결정한다. 이때 부분패턴의 길이 $[b]$ 는 $2 \leq b \leq \ell$ 인 조건을 만족하도록 결정된다. 도 4에서와 같이 길이가 L 인 패턴(41)은 길이가 b 인 연속된 부분패턴을 $L-b+1$ 개 포함하게 되므로, 부분패턴의 수는 자동으로 결정된다.

<22> 패턴집합에서 부분패턴을 일관된 방법으로 추출하기 위하여 부분패턴의 길이 $[b]$ 를 $2 \leq b \leq \ell$ 조건에서 임의로 정하고 각 패턴에서 부분패턴을 추출한다. 패턴에서 추출하는 부분패턴의 길이를 고정시키면 서로 다른 길이를 갖는 패턴에 대하여 추출할 수 있는 부분패턴의 개수가 다르게 된다. 본 발명에서는 모든 패턴에 대하여 최소 길이를 갖는 최소패턴에서 생성할 수 있는 수 만큼의 길이 b 인 부분패턴만을 추출한다. 왜냐하면 최소길이 패

턴의 부분패턴의 개수[n]보다 많은 부분패턴을 추출하여 이용하더라도 이동거리를 n이상 증가시킬 수 없기 때문이다. 도 1에서 생성한 4개의 부분패턴(52)에서 최소패턴의 n이 2라서 2개의 연속된 부분패턴을 선택해야 할 경우 (53), (54), (55)의 3가지 방법이 존재하는 것을 도 5에 예시하였다.

<23> 한편 추출한 부분패턴만을 이용하여 패턴매칭을 할 경우 그 결과가 부정확할 수 있다. 따라서 정확한 전체 매칭을 위해 각 부분패턴에 대한 메타정보를 부분패턴 테이블에 저장(34)한다. 메타정보는 부분패턴을 추출한 모패턴의 길이에 대한 정보와, 그 패턴에서 추출한 부분패턴의 위치에 대한 정보를 포함한다. 예를 들어, 도면 5에서 부분패턴(54)의 경우, 처음 부분패턴 (541)에 대해서 패턴의 길이정보 6과 패턴에서 부분패턴의 위치정보 -1을, 두 번째 부분패턴 (542)에 패턴의 길이정보 6과 위치정보 -2를 부분패턴 테이블에 저장한다.

<24> 서로 다른 모패턴에서 추출된 부분패턴들이 동일할 수 있다. 이때 중복되는 부분패턴들을 모두 부분패턴 테이블에 별도로 저장하는 것은 비효율적이므로 부분패턴 단위로 합병하여 부분패턴 테이블을 구성(35)한다. 예를 들어, 도 6과 같이 패턴 (61)과 (62)일 때, 각 패턴에는 (612)와 (621)과 같이 중복된 부분패턴이 존재한다. 중복된 부분패턴은 (63)과 같이 합병하여 테이블로 구성한다.

<25> (B) 매칭확인단계

<26> 상기에서 추출된 부분패턴 집합을 이용하여 패턴매칭하는 과정을 도 7에 도시하였다.

<27> 먼저, 페이로드 또는 문자열(이하 단순히 '페이로드'라 함)의 처음부터 패턴매칭을 진행한다(71). 현재 위치가 페이로드의 끝 오프셋인지 비교(72)한다. 현재 오프셋이 끝이 아니라면 부분패턴의 길이[b]만큼 문자열을 선택(73)하여 1단계로 일치하는 부분패턴이 존재하는지 확인한다(74). 만약 일치하는 부분패턴이 존재하지 않으면 페이로드의 현재 오프셋을 부분패턴의 개수[n] 만큼 이동(76)하고 오프셋 비교(72)부분으로 돌아가 반복한다. 일치하는 부분패턴이 존재하면 2단계로 부분패턴 테이블에 저장된 메타정보인 모패턴의 길이정보와 부분패턴의 위치정보를 이용하여 모패턴의 전체 문자열과 비교(77)한다. 문자열이 모두 일치하는 패턴을 찾게 되면 패턴매칭 결과를 보이고(79) 종료한다. 이 과정을 패킷이 들어올 때 마다 수행하여 알려진 규칙에 대한 패턴을 찾게 된다.

<28> 한편 부분패턴 테이블에 저장된 부분패턴들을 페이로드에서 선택된 문자열들과 효과적으로 비교하고 부분패턴을 저장하는데 필요한 공간을 줄이기 위하여 부분패턴을 해시함수에 의한 해시 값으로 대체할 수 있다. 이를 위하여 상기 단계 (A)에는, (d) 추출된 부분패턴에 대한 해시값을 정하는 소단계;를 추가로 포함하며, 상기 단계 (B)의 (a)에서는 비교부분에 대한 해시값을 구하고 상기 부분패턴의 해시값들과의 일치여부를 비교하도록 한다.

<29> 이러한 방법 채택시, 부분패턴에 대한 해시 값은 부분패턴 테이블의 인덱스로 사용될 수 있으므로 부분패턴들이 부분패턴 테이블에 실제 저장되지는 않는다.

도면의 간단한 설명

<30> 도 1은 패턴의 길이가 6바이트인 패턴에서 3바이트 크기를 갖는 부분패턴 4개와 4바이트 크기를 갖는 부분패턴 3개를 예로 보여주는 도면이다.

<31> 도 2는 패턴에서 추출한 4개의 부분패턴을 이용하여 페이로드에서 패턴의 일치 여부를 확인 없이 패턴의 개수 4만큼 건너뛰며 패턴매칭하는 예를 보여주는 도면이다.

<32> 도 3은 본 발명에서 패턴에서 부분패턴을 추출하는 과정을 순서도로 보여주는 도면이다.

<33> 도 4는 본 발명에서 다중 패턴매칭을 위해 추출할 부분패턴과 패턴과의 관계를 나타내는 도면이다.

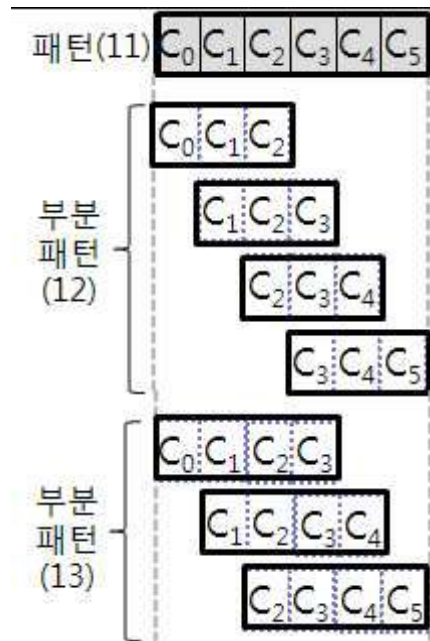
<34> 도 5는 본 발명에서 다중 패턴매칭을 위해 패턴에서 추출한 부분패턴 중에서 필요한 부분패턴을 선택하는 실패를 보여주는 도면이다.

<35> 도 6은 본 발명에서 다중 패턴매칭을 위해 패턴에서 추출한 부분패턴 중에서 중복되어 나타나는 부분패턴을 합병하는 실패를 보여주는 도면이다.

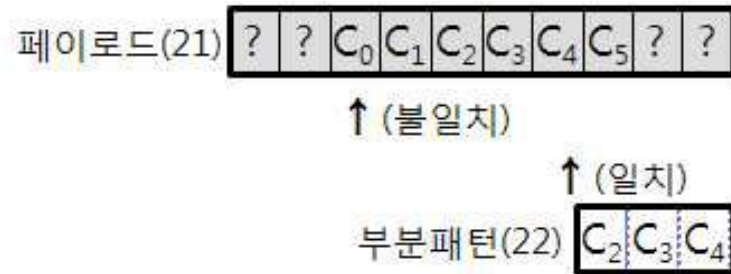
<36> 도 7은 본 발명에서 여러 개의 연속된 부분패턴을 이용하여 페이로드의 처음부터 끝까지 스캐닝하며 패턴을 검색하는 과정을 순서도로 보여주는 도면이다.

도면

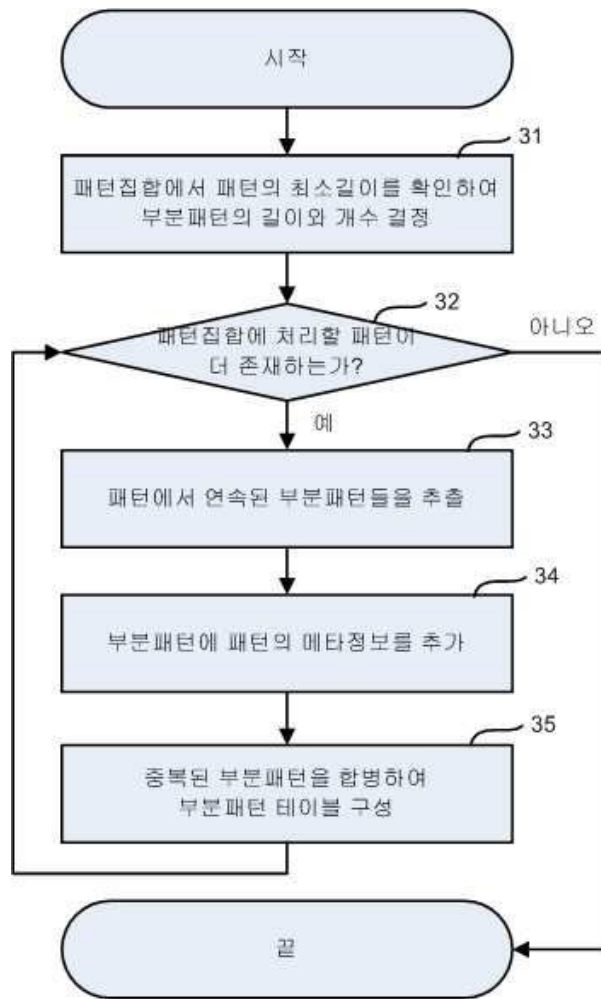
도면1



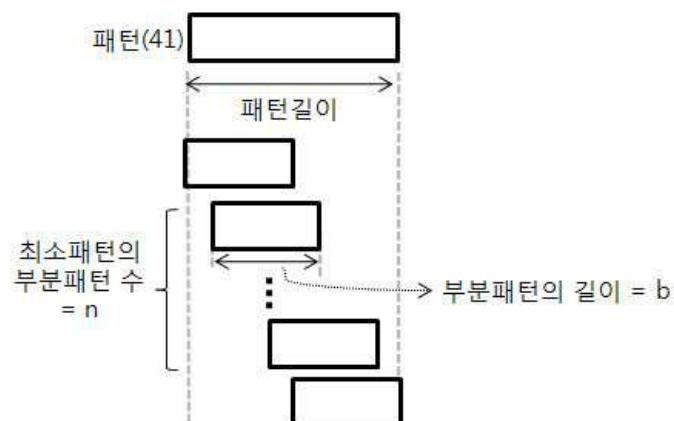
도면2



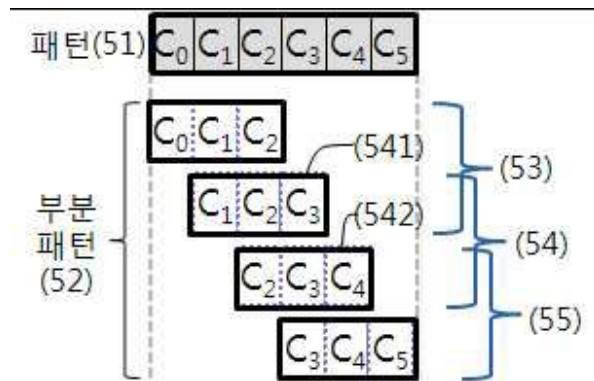
도면3



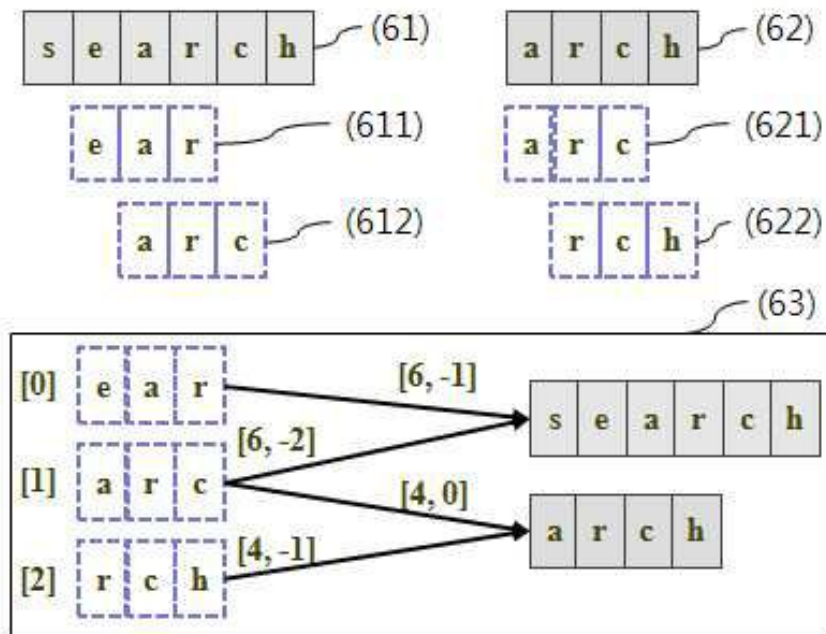
도면4



도면5



도면6



도면7

