

화상회의용 워터마크 기술(技術) 보고서

김동현(Kim DongHyeon, wlswudpdf31@kookmin.ac.kr)

November 18, 2024

요약

본 보고서는 화상회의용 워터마크 기술, 카멜레온을 소개(?)한다. 기존 화상회의 서비스가 제공하는 워터마크 기술은 유출자를 식별할 수 없고, 워터마크 진위성을 파악할 수 없으며, 유출자가 워터마크를 쉽게 지울 수 있다. 카멜레온은 이 한계점을 극복하기 위해 개발한 새로운 워터마크 기술이다. 카멜레온은 텍스트 워터마크와 QR코드 워터마크를 화면에 반복 투사(?)하여 워터마크가 일부 지워지더라도 추출할 수 있다. 텍스트 워터마크는 화상회의 영상 내 텍스트와 구분하기 어렵게 만들었고, QR 코드는 화상회의 색상과 비슷하게 생성하여, AI가 워터마크를 영상과 구분하지 못하도록 하였다.

목차

1 서론	3
2 기술 설명	4
2.1 텍스트 워터마크	4
2.1.1 메시지 구성	5
2.1.2 삽입 방식	5
2.2 QR코드 워터마크	6
2.2.1 메시지 구성	6
2.2.2 삽입 방식	6
2.3 워터마킹 삽입 및 추출 시나리오	7
3 워터마크 추출 검증	7
3.1 워터마크 훼손 없는 경우	9
3.2 AI로 인해 워터마크 훼손 있는 경우	9
3.3 영상 워터마크 제거 도구 사용 후	10
4 토의	10
5 결론	10

1 서론

오늘날 많은 기업이 화상회의 서비스를 사용하고 있다. 코로나 19 이후 기업은 재택근무, 원격근무를 추진했다. 많은 직원이 화상회의 서비스 사용을 추구하는 상황에서, 기업은 화상회의와 관련한 보안 이슈를 고려할 필요가 있다. 이미 화상회의 서비스를 제공하는 기업에서도 보안을 강화하고 있고, 사용하는 기업에서도 화상회의를 안전하게 진행하기 위해 노력하고 있다. 담긴 회의 영상이 유출되면, 기업은 큰 피해를 입을 수 있다. 외부 공격자가 기업 내부로 들어와 회의 영상을 탈취하여 유출하거나, 회의 참여자가 자신의 이익을 위해 타 기업에 영상을 유출할 수도 있다. 따라서 기업은 화상회의 영상 유출로 인한 피해를 막기 위해 노력해야 한다.

디지털 워터마크는 이미지, 동영상, 오디오, 문서와 같은 디지털 콘텐츠에 삽입하는 정보 또는 코드이다. 콘텐츠가 유출됐을 때 워터마크를 이용하면, 유출 사고 피해를 줄일 수 있다. 대표적으로, 워터마크를 이용하여 콘텐츠의 소유권을 보호할 수 있다. 화상회의 화면에 호스트 정보를 워터마크로 삽입하면, 누군가 화상회의를 기록하여 영상을 유출했을 때, 워터마크를 추출하여 이 영상이 호스트의 것임을 증명할 수 있다. Zoom은 이미 워터마크 기술을 이용자에게 제공하고 있다. Zoom은 호스트 이메일을 화면에 띠워 모든 참석자가 이메일을 볼 수 있게 한다. 그림 1은 Zoom에서 워터마크 기능을 사용했을 때 보이는 화상회의 화면이다.

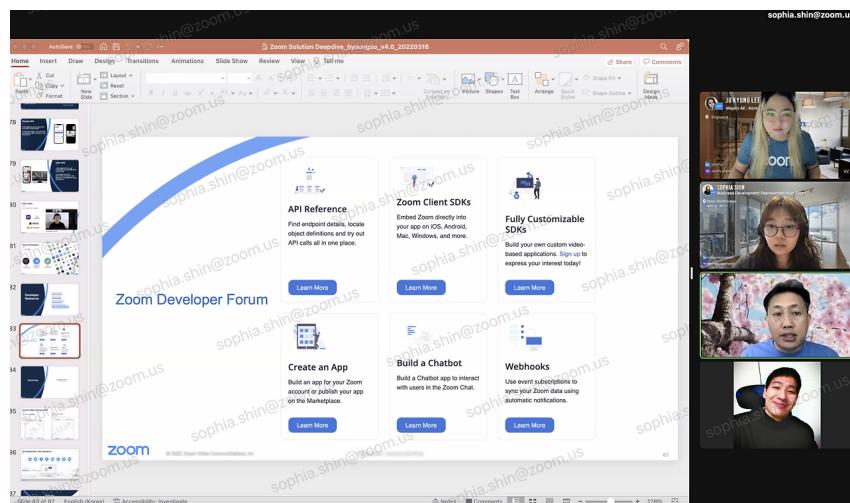


그림 1: Zoom에서 워터마크를 삽입한 화상회의 화면

이 외에도 워터마크 기술을 제공하는 화상회의 서비스는 대부분 화상회의 호스트의 식별정보를 워터마크 정보로 사용한다. 회의 참석자 혹은 제 3자가 콘텐츠를 유출하면, 콘텐츠로부터 워터마크를 추출하여 콘텐츠의 소유권이 호스트에 있음을 증명할 수 있다. 그러나 기존 워터마크 기술에는 한계점이 있다.

- **유출자 식별 불가능.** 일반적으로 워터마크에는 회의영상 소유자 정보만 사용한다. 이 경우 영상의 소유권을 주장할 수 있지만, 누구로부터 유출됐는지 파악할 수 없다. 유출자를 알 수 없으면, 영상 유출에 대해 누구에게도 책임을 물을 수 없고, 향후 회의영상 유출에 대해서도 대비할 수 없다.
- **워터마크 진위성 확인 불가능.** 유출자를 식별하기 위해 호스트 정보뿐만 아니라 회의 참석자 정보도 워터마크로 삽입하더라도, 유출 영상의 워터마크가 진짜 메시지를 담고 있는지는 알 수 없다. 만약 유출

자가 자신의 워터마크를 지우고, 다른 사람의 정보로 워터마크를 삽입한다면, 다른 유출자를 색출하게 된다.

- **워터마크 삭제 가능.** 원본 영상의 손상 없이 워터마크를 훼손하면, 워터마크는 그 효력을 잃는다. AI를 기반으로한 워터마크 제거 도구는 전문지식 없이도 텍스트 워터마크를 쉽게 지울 수 있다.

카멜레온은 이런 한계를 극복하고자 개발한 워터마크 기술이다. 카멜레온은 워터마크 메시지에 호스트 식별 정보 뿐만 아니라 회의 참석자 식별정보를 추가했다. 참석자 중 누군가가 회의를 녹화하여 유출하면, 참석자 식별정보를 확인하여 유출자를 알 수 있다. 카멜레온은 이 식별정보와 호스트 키를 사용하여 메시지인증코드를 생성하고, 워터마크 메시지의 진위성을 확인할 수 있다. 카멜레온은 워터마크를 텍스트와 QR코드 형태로 생성한다. 특히 QR코드 형태의 워터마크는 AI가 워터마크를 지울 수 없도록 특정 방법으로 색상을 변경했다. 본 보고서는 카멜레온의 워터마크 생성 과정을 자세히 설명하고, 워터마크 추출 결과를 분석한다.

보고서 구조는 다음과 같다. 2절은 카멜레온이 워터마크 메시지로 무엇을 사용했는지, 워터마크를 어떤 방식으로 삽입했느지 등을 설명한다. 3절은 카멜레온으로 콘텐츠에 워터마크를 삽입하고, 콘텐츠로부터 워터마크가 잘 추출되는지 검증한다. 일반적인 화상회의 영상으로부터 추출한 결과와 AI 제거 도구로 워터마크를 손상한 후 추출한 결과를 제시한다. 4절은 카멜레온의 한계점이나 개선점 등을 토의한다. 5절은 결론으로 보고서를 마친다.

2 기술 설명

카멜레온은 기존 화상회의용 워터마크 기술이 가지는 한계를 극복한 워터마크 기술이다. 카멜레온은 회의 호스트 정보 뿐만 아니라 회의 참석자 정보를 추가하여, 회의영상을 녹화하는 참석자가 누구인지 알 수 있다. 또한 워터마크 메시지에 MAC을 삽입하여 메시지의 진위성을 확인 할 수 있다. 카멜레온은 기존 워터마크 기술과 비슷한 방식으로 텍스트를 삽입했고, QR코드를 이용하여 그림 형태 워터마크를 삽입했다. QR코드 워터마크는 회의영상을 실시간으로 캡처하여 QR코드 색을 캡처한 사진의 색과 비슷하게 생성한다. 이를 통해 AI가 QR코드와 회의영상과 구분할 수 없게 했다.

2.1 텍스트 워터마크

텍스트 워터마크는 콘텐츠 이용자가 메시지를 그대로 볼 수 있는 워터마크이다. 텍스트 워터마크를 사용하여 회의 참석자가 회의를 녹화하고 유출하면 안된다는 경각심을 가질 수 있다. 따라서 기존 영상 화상회의 서비스가 제공하는 워터마크 기술과 비슷하게 텍스트 워터마크를 삽입했다. 그러나 기존 워터마크는 유출자를 식별하지 못하고, 진위성을 판별할 수 없다. 카멜레온은 기존 텍스트 워터마크 메시지에 유출자 정보와 MAC을 삽입하여, 이 문제를 해결했다. 그리고 AI 제거 도구가 텍스트 워터마크를 지우지 못하는 경우를 고려하여 텍스트의 형태를 설정했다.

2.1.1 메시지 구성

카멜레온은 워터마크 메시지로 주의 사항, 호스트 식별 정보, 참석자 식별 정보, 현재 시간, 메시지 인증 코드를 사용한다.

- **주의사항.** 텍스트 워터마크는 회의 참석자에게 하고 싶은 말을 전달할 수 있다. 참석자에게 화상회의 를 녹화하지 말라는 주의사항을 전달하여, 실수로 녹화하는 일이 없도록 한다. 카멜레온은 ”Do Not Copy!”를 메시지로 삽입했으며, 다른 메시지를 사용할 수 있다.
- **호스트 식별정보.** 화상회의 영상의 소유권은 호스트에게 있다. 회의 영상이 유출됐을 때, 워터마크로 부터 호스트 정보를 얻어 영상 소유권이 호스트에게 있음을 증명할 수 있다.
- **참석자 식별정보.** 회의 참석자 식별 정보는 각 참석자 PC마다 서로 다르다. 회의영상이 유출되었을 때, 참석자 식별 정보를 확인하여 어느 참석자의 PC에서 유출되었는지 확인할 수 있다.
- **현재시간.** 회의를 진행하고 있는 현재 시간을 워터마크 삽입하여, 유출된 회의가 언제 진행한 회의인지 알 수 있다. 현재 시간은 실시간으로 갱신하여 새로운 메시지를 삽입한다.
- **메시지 인증 코드.** 메시지 인증 코드(이하 MAC)는 데이터가 변조되었는지 검증 할 수 있도록 데이터 뒤에 덧붙이는 코드이다. 코드를 생성할 때는 호스트의 키를 사용하여, 호스트만 데이터 변조 여부를 알 수 있다. 카멜레온은 호스트 식별 정보, 참석자 식별 정보, 현재 시간 세 개의 메시지에 대해 MAC값을 계산한다. 회의영상이 유출되면, 호스트는 자신의 키를 사용하여 워터마크 메시지의 MAC을 계산하고, 워터마크로 삽입된 MAC과 같은지 확인하여 메시지 진위성을 확인한다.

그림 2은 텍스트 워터마크를 삽입한 회의영상 사진 일부이다. 주의사항으로 “Do Not Copy!”를 보여주어, 회의 참석자들에게 회의영상을 복사하지 말라고 권고한다. Alice는 회의 호스트, Bob은 회의 참석자 중 한명의 식별정보이다. Bob이 아닌 다른 참석자의 회의 영상에서는 해당 부분이 다른 정보로 나타난다. 그리고 현재 시간 2024...가 나타나 있으며, 밑에는 Alice-Bob-현재시간에 대한 MAC 값을 삽입했다. 사용한 키는 호스트 Alice의 키이며, 여기서는 ‘chameleon’를 키로 사용하고, 알고리즘은 SHA256을 사용했다.

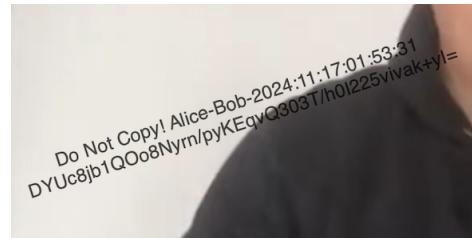


그림 2: 텍스트 워터마크 구성

2.1.2 삽입 방식

다양한 형태로 텍스트를 삽입해본 결과, 텍스트 워터마크가 회의영상 속 텍스트와 유사하면, AI 제거 도구는 텍스트 워터마크를 지우지 못한다. 혹은 지우더라도 원본을 크게 훼손한다. 따라서 폰트, 글자 크기, 글자 색 등을 결정할 때, 화상회의 화면에 주로 보이는 텍스트와 유사하게 구성했다. 일반적으로 화상회의에서는 문서를 보여주거나 발표자료를 보여주기 때문에, 이 때 자주 사용하는 글자 형태에 맞췄다. 그러나 이렇게 텍스트 워터마크를 삽입하면, 회의 참석자 또한 워터마크와 문서 속 텍스트와 구분하기 어려워 회의 참여가 불편할 수 있다. 따라서 텍스트 워터마크를 기울여 문서 내 글과 구분했다.

텍스트 워터마크는 많이 삽입할수록 추출할 가능성이 높다. 화면을 가리지 않으려고 일부에만 워터마크를 삽입할 경우, 유출자는 해당 부분을 자르고 영상을 유출시킬 수 있다. 따라서 카멜레온은 텍스트 워터마크를 화면 전체에 골고루 삽입했다. 또한 카멜레온은 워터마크에 애니메이션 효과를 줬다. AI 워터마크 제거 도구는 화면의 특정 영역을 선택해 워터마크를 지우는 기능이 있다. 워터마크의 위치가 고정되면, 선택 영역이 좁아져 원본을 크게 훼손하지 않고 워터마크를 지울 수 있다. 워터마크 위치를 옮기면 선택 영역이 넓어진다. 또한 워터마크가 움직이기 때문에 화면을 계속 가리지 않는다.

2.2 QR코드 워터마크

QR코드는 2차원 매트릭스 바코드의 일종이다. QR 코드는 오류정정기법을 사용해, 코드를 일부 훼손하더라도 그 안에 있는 메시지를 유지할 수 있다. 워터마크는 콘텐츠가 시간이 지남에 따라 열화되면서 훼손될 수 있고, 유출자에 의해 훼손될 수도 있다. QR 코드를 이용하여 워터마크를 삽입하면, 이러한 훼손에도 워터마크 내 메시지를 추출할 수 있다. 카멜레온은 QR 코드의 특성을 활용하기 위해, QR코드로 워터마크를 삽입한다.

2.2.1 메시지 구성

QR코드 워터마크가 담는 메시지는 텍스트 워터마크와 동일하다. 주의사항을 제외한 참여자 식별정보, 유출자 식별정보, 현재시간, MAC값을 담는다. 다만, 4 가지 정보를 하나의 QR코드에 담지 않고, MAC값은 다른 QR코드로 만들어, 두 개의 QR코드를 만든다. QR코드 크기는 78px로, 44 바이트 메시지를 담을 수 있다.

그림 3은 카멜레온이 QR코드 워터마크를 삽입한 영상 일부이다. 왼쪽 위 QR코드와 오른쪽 QR 코드가 다른 것을 알 수 있는데, 각각 식별정보와 현재시간을 담은 QR코드, MAC값을 담은 QR 코드이다.



그림 3: QR코드 워터마크 구성

2.2.2 삽입 방식

카멜레온은 QR코드를 삽입 하기 전 QR 코드 색상을 결정한다.

색상은 화상회의 화면에 의존한다. 화상회의 화면을 캡처한 후 QR 코드를 삽입할 위치에 해당하는 픽셀의 평균 RGB를 계산한다. 이 값을 QR코드 색으로 결정한다. 그림 4는 QR코드 워터마크 색상 변경 방식을 그림으로 표현한 것이다.

이 방식은 두 가지 기대효과를 가진다. AI 워터마크 제거 도구는 QR코드 워터마크를 지울 수 있다. QR코드 색상을 화상회의 화면 색과 유사하게 변경하면 AI는 QR코드 워터마크와 화상회의 영상을 구분하지 못하기를 구분할 수 있다. 실험 결과, QR코드를 흑백으로 삽입했을 때보다, 색상을 비슷하게 만들어 삽입했을 때 AI는 덜 훼손하는 것을 확인했다. 그리고 QR코드를 화면 전체에 투사하면 회의 참석자는 QR코드에 의해 화면을 잘 보지 못할 수 있다. 따라서 QR코드 워터마크의 투명도를 높여 참석자가 회의영상을 시청할 때 불편함을



그림 4: QR코드 워터마크 색상 변경 방식

줄일 수 있다. QR코드 색상을 뒷 배경 색과 비슷하게 만들면, QR코드의 투명도가 높아지는 효과를 줄 수 있다.

카멜레온은 두 종류의 QR코드를 화면에 체크무늬 형태로 출력한다. QR코드를 화면에 가득 채우면 회의에 방해가 될 수 있고, 향후에 비어있는 공간을 활용하여 다른 형태의 워터마크를 삽입할 수 있게 했다.

2.3 워터마킹 삽입 및 추출 시나리오

(미완성입니다.)

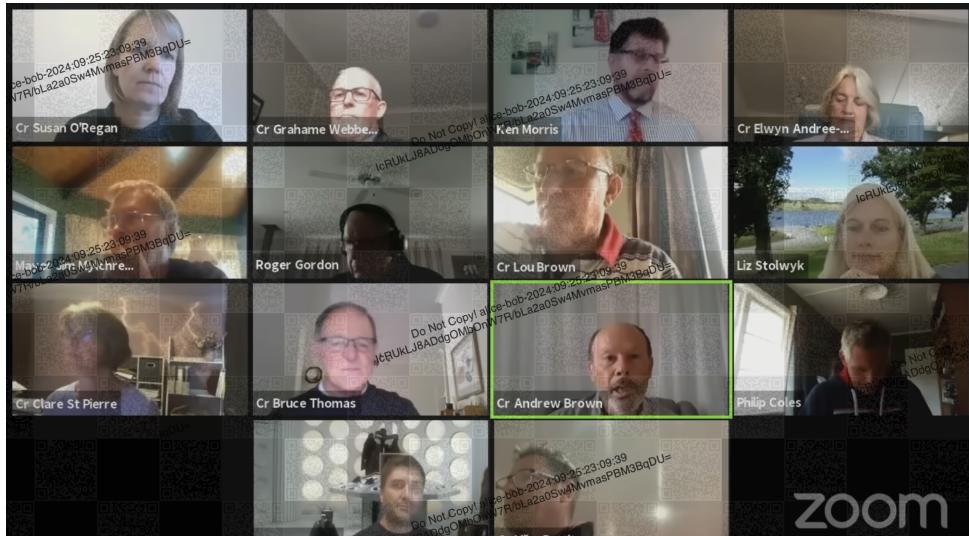


그림 5: 카멜레온 워터마크 삽입 화면

3 워터마크 추출 검증

(미완성입니다.)

- Both: 텍스트 워터마크와 QR코드 워터마크 둘 다 추출 가능.
- Text: 텍스트 워터마크만 추출 가능.
- QR: QR코드 워터마크만 추출 가능.
- None: 워터마크 추출 불가능.

P_i, T_i 사진 제시 필요.

3.1 워터마크 훼손 없는 경우

	P_1	P_2	P_3	T_1	T_2	T_3	T_4
기본	Both						
크기 70% 감소	Both						
크기 30% 감소	Text						
밝기 30% 감소	Both						
밝기 70% 감소	Both						
채도 50% 감소	Both						
흑백	Both						
색상반전	Both						

Table 1: AI 적용 전

(표 결과 분석, 표 디자인 다른 것도 고려해보기)

3.2 AI로 인해 워터마크 훼손 있는 경우

	P_1	P_2	P_3	T_1	T_2	T_3	T_4
기본	QR	QR	QR	Text	Text	Text	Text
크기 70% 감소	QR	QR	QR	Text	Text	Text	Text
크기 30% 감소	None	None	None	Text	Text	Text	Text
밝기 30% 감소	QR	QR	QR	Text	Text	Text	Text
밝기 70% 감소	QR	QR	QR	Text	Text	Text	Text
채도 50% 감소	QR	QR	QR	Text	Text	Text	Text
흑백	QR	QR	QR	Text	Text	Text	Text
색상반전	QR	QR	QR	Text	Text	Text	Text

Table 2: AI 적용 후

(표 결과 분석)

3.3 영상 워터마크 제거 도구 사용 후

(영상 워터마크 제거 사진 첨부)

4 토의

(아직 미완성입니다. '토의' 절에는 카멜레온 기술 설명에 없는 내용에 대해 논의합니다.)

- **텍스트 투명도.** 텍스트는 불투명도가 높아 회의에 방해 될수 있다. 낮은 보안을 요구하는 상황에서는 불투명도를 낮출 수 있다.
- **QR 코드 크기.** QR 코드 크기를 늘리면 메시지 크기를 늘릴 수 있다. 또한 QR코드 크기가 다양하면 AI가 쉽게 못 지울 수 있다.

5 결론

(미완성입니다.)